

[illegible]

METHODS TO IDENTIFY COMPOUNDS FOR DISRUPTING PROTEIN/PROTEIN INTERACTIONS

Background of the Invention

The present invention relates to a novel method to identify inhibitors of protein/protein interactions.

Background

Modulation of protein/protein interactions is an attractive target for drug discovery and development. Potential methods by which drugs can regulate protein/protein interactions are numerous, including, for example, regulation of expression of one or more of the binding proteins, modulation of post-translational modification, and direct interference with the capacity of one protein to bind to one or more binding partners. More importantly, recent observations make it increasingly clear that supramolecular protein complexes, involving two or more binding proteins, play an important and essential roles in signal transduction, gene expression, cell proliferation and duplication, and cell cycle progression. For example, in the repair of UV damaged DNA, a so-called "repairosome" that contains over ten individual proteins is assembled into a complex which can then carry out the necessary repair. Likewise, gene transcription occurs through the concerted action of greater than twenty proteins. Signal transduction proteins, such as receptor protein kinases, are part of large complexes with many proteins. Contacts through *Src* homology type 2 (SH2) domains on the receptor kinases, for example, are noteworthy protein interaction which are part of one or more enzymatic cascade important for many metabolic processes. Disrupting the binding capacity of one or more proteins which form any of these larger complex is therefore an important and untapped method to control action of the overall complex.

Protein/protein interactions have been discovered and characterized by a variety of methods: (i) standard biochemical affinity

- 2 -

methods such as chromatography or co-immunoprecipitations; (ii) gel overlay methods; (iii) co-purification by traditional biochemistry; and (iv) two-hybrid analysis [Fields and Song, *Nature* 340:245-246 (1989); Fields, *Methods: A Companion to Methods in Enzymology* 5:116-124 (1993); U.S. Patent 5,283, 173 issued February 1, 1994 to Fields, *et al.*]. The most recent of these approaches, the two hybrid method, has enjoyed broad application because of its relative ease of use for gene identification from cDNA fusion libraries. [See Chien *et al.*, *Proc. Natl. Acad. Sci. (USA)* 88:9578-9582 (1991); Dallon and Treisman, *Cell* 72:223-232 (1993); and Durfee, *et al.*, *Genes and Devol.* 7:555-569 (1993)].

The two hybrid system is based on targeting and identifying a protein/protein interaction through the use of a reporter system. The described two hybrid systems either use the yeast Gal4 DNA binding domain or the *E. coli* *lexA* DNA binding domain and couple this region to a transcriptional activator such as Gal4 or VP16 that drives a reporter like β galactosidase or HIS3.

In principle the two hybrid assay could be used for drug screening. [See WO 96/03501 and WO 96/03499.] In such a scenario, loss of β galactosidase or HIS3 activity would be identified after the yeast strain is treated with a compound. In practice, however, use of the two hybrid system is technically undesirable for several reasons. In instances where the β galactosidase or HIS3 protein are employed as the reporter protein, a loss of activity is particularly difficult to detect because the expressed reporter protein is too long lived to be used in a high throughput mode. If a candidate binding inhibitor compound is metabolized faster than the previously expressed reporter protein is turned over, it is difficult to detect inhibitory action of the candidate drug while a reporter protein is still active. In high throughput screening, the loss of a positive signal, for example, β galactosidase or HIS3 is impossible to detect. Present robotized screening and detection methods are simply not sufficiently sensitive or robust to detect loss of a signal.

- 3 -

Thus there is a need in the art to develop a rapid screening method that gives a positive signal, as opposed to a negative signal, when a protein/protein interaction is disrupted. Such a system must be capable of using protein interactions that are initially detected by any of the above mentioned approaches and must be sufficiently robust to detect a gain of function when a protein interaction is lost. In essence, the screening method must give a signal when an interaction is lost, not lose a signal when an interaction is lost. Such a system must be sensitive to subtle interactions, in particular ones that are caused by post-translational modification like protein phosphorylation. Finally for large scale screening, such as high throughput screening, the system must be manipulable such that a large signal-to-noise ratio can be easily detected.

Brief Summary of the Invention

In one aspect, the present invention provides materials that are useful for the identification of compounds which inhibit interaction between known binding partner proteins. See Figure 1. The invention provides host cells transformed or transfected with DNA comprising: (i) a repressor gene encoding DNA binding protein that acts as a repressor protein, said repressor gene under transcriptional control of a promoter; (ii) a selectable marker gene encoding a selectable marker protein; said selectable marker gene under transcriptional control of an operator; said operator regulated by interaction with said repressor protein; (iii) a first recombinant fusion protein gene encoding a first binding protein or binding fragment thereof in frame with either a DNA binding domain of a transcriptional activating protein or a transactivating domain of a transcriptional activating protein; and (iv) a second recombinant fusion protein gene encoding a second binding protein or binding fragment thereof in frame with either a DNA binding domain of a transcriptional activating protein or a transactivating domain of a transcriptional activating protein, whichever domain is not encoded by the first

- 4 -

fusion protein gene, said second binding protein or binding fragment thereof capable of interacting with said first binding protein or binding fragment thereof such that interaction of said second binding protein or binding fragment thereof and said first binding protein or binding fragment thereof brings into proximity a DNA binding domain and a transactivating domain forming a functional transcriptional activating protein; said functional transcriptional activating protein acting on said promoter to increase expression of said repressor gene.

The invention comprehends host cells wherein the various genes and regulatory sequences are encoded on a single DNA molecule as well as host cells wherein one or more of the repressor gene, the selectable marker gene, the first recombinant fusion protein gene, and the second recombinant fusion protein gene are encoded on distinct DNA expression constructs. In a preferred embodiment, the host cells are transformed or transfected with DNA encoding the repressor gene, the selectable marker gene, the first recombinant fusion protein gene, and the second recombinant fusion protein gene, each encoded on a distinct expression construct. Regardless of the number of DNA expression constructs introduced, each transformed or transfected DNA expression construct further comprises a selectable marker gene sequence, the expression of which is used to confirm that transfection or transformation was, in fact, accomplished. Selectable marker genes encoded on individually transformed or transfected DNA expression constructs are distinguishable from the selectable marker gene regulated by the *tet* operator in that expression of the selectable marker gene regulated by the *tet* operator is central to the preferred embodiment; *i.e.*, regulated expression of the selectable marker gene by the *tet* operator provides a measurable phenotypic change in the host cell that is used to identify a binding protein inhibitor. Selectable marker genes encoded on individually transformed or transfected DNA expression constructs are provided as determinants of successful transfection or transformation of the individual

- 5 -

DNA expression constructs. Preferred host cells of the invention include transformed *S. cerevisiae* strains designated Y1596 and Y1584 which were deposited August 13, 1996 with the American Type Culture Collection (ATCC), 12301 Parklawn Drive, Rockville, Maryland 20852, and assigned Accession Numbers ATCC 74384 and ATCC 74385, respectively.

The host cells of the invention include any cell type capable of expressing the heterologous proteins required as described above and which are capable of being transformed or transfected with functional promoter and operator sequences which regulate expression of the heterologous proteins also as described. In a preferred embodiment, the host cells are of either mammalian, insect or yeast origin. Presently, the most preferred host cell is a yeast cell. The preferred yeast cells of the invention can be selected from various strains, including the *S. cerevisiae* yeast transformants described in Table 1. Alternative yeast specimens include *S. pombe*, *K. lactis*, *P. pastoris*, *S. carlsbergensis* and *C. albicans*. Preferred mammalian host cells of the invention include Chinese hamster ovary (CHO), COS, HeLa, 3T3, CV1, LTK, 293T3, Rat1, PC12 or any other transfectable cell line of human or rodent origin. Preferred insect cells lines include Sf9 cells.

In a preferred embodiment, the selectable marker gene is regulated by an operator and encodes an enzyme in a pathway for synthesis of a nutritional requirement for said host cell such that expression of said selectable marker protein is required for growth of said host cell on media lacking said nutritional requirement. Thus, as in a preferred embodiment where a repressor protein interacts with the operator, transcription of the selectable marker gene is down-regulated and the host cells are identified by an inability to grow on media lacking the nutritional requirement and an ability to grow on media containing the nutritional requirement. In a most preferred embodiment, the selectable marker gene encodes the HIS3 protein, and host cells transformed or transfected with a HIS3-encoding DNA expression construct are selected following growth on media in the presence

- 6 -

and absence of histidine. The invention, however, comprehends any of a number of alternative selectable marker genes regulated by an operator. Gene alternatives include, for example *URA3*, *LEU2*, *LYS2* or those encoding any of the multitude of enzymes required in various pathways for production of a nutritional requirement which can be definitively excluded from the media of growth. In addition, conventional reporter genes such as chloramphenicol acetyltransferase (CAT), firefly luciferase, β -galactosidase (β -gal), secreted alkaline phosphatase (SEAP), green fluorescent protein (GFP), human growth hormone (hGH), β -glucuronidase, neomycin, hygromycin, thymidine kinase (TK) and the like may be utilized in the invention.

In the preferred embodiment, the host cells include a repressor protein gene encoding the tetracycline resistance protein which acts on the *tet* operator to decrease expression of the selectable marker gene. The invention, however, also encompasses alternatives to the *tet* repressor and operator, for example, *E. coli trp* repressor and operator, *his* repressor and operator, and *lac* operon repressor and operator.

The DNA binding domain and transactivating domain components of the fusion protein may be derived from the same transcription factor or from different transcription factors as long as bringing the two domains into proximity permits formation of a functional transcriptional activity protein that increases expression of the repressor protein with high efficiency. A high efficiency transcriptional activating protein is defined as having both a DNA binding domain exhibiting high affinity binding for the recognized promoter sequence and a transactivating domain having high affinity binding for transcriptional machinery proteins required to express repressor gene mRNA. The DNA binding domain component of a fusion protein of the invention can be derived from any of a number of different proteins including, for example, LexA or Gal4. Similarly, the transactivating component of the invention's fusion proteins can be derived from a number of different transcriptional activating proteins, including for example, Gal4 or

- 7 -

VP16. In one embodiment of the invention, polynucleotides encoding binding partner proteins CREB and CBD are inserted in plasmids pVP16-CREB and pLexA-CBD, respectively, which were deposited with the ATCC and assigned Accession Numbers ATCC 98138 and ATCC 98139, respectively.

The promoter sequence of the invention which regulates transcription of the repressor protein can be any sequence capable of driving transcription in the chosen host cell. The promoter may be a DNA sequence specifically recognized by the chosen DNA binding domain of the invention, or any other DNA sequence with which the DNA binding domain of the fusion protein is capable of high affinity interaction. In a preferred embodiment of the invention, the promoter sequence of the invention is either a HIS3 or alcohol dehydrogenase (ADH) promoter. In a presently most preferred embodiment, the ADH promoter is employed in the invention. The invention, however, encompasses numerous alternative promoters, including, for example, those derived from genes encoding HIS3, ADH, URA3, LEU2 and the like.

In another aspect, the invention provides methods to identify molecules that inhibit interaction between known binding partner proteins. In one embodiment, the invention provides a method to identify an inhibitor of binding between a first binding protein or binding fragment thereof and a second binding protein or binding fragment thereof comprising the steps of (a) growing host cells transformed or transfected as described above in the absence of a test compound and under conditions which permit expression of said first binding protein or binding fragment thereof and said second binding protein or binding fragment thereof such that said first binding protein or fragment thereof and second binding protein or binding fragment thereof interact bringing into proximity said DNA binding domain and said transactivating domain forming a functional transcriptional activating protein; the transcriptional activating protein acting on said promoter to increase

- 8 -

expression of said repressor protein; said repressor protein interacting with said operator such that said selectable marker protein is not expressed; (b) confirming lack of expression of said selectable marker protein in said host cell; (c) growing said host cells in the presence of a test compound; and (d) comparing expression of said selectable marker protein in the presence and absence of said test compound wherein increased expression of said selectable marker protein is indicative that the test compound is an inhibitor of binding between said first binding protein or binding fragment thereof and said second binding protein or binding fragment thereof.

In a most preferred embodiment, the invention provides a method to identify an inhibitor of binding between a first binding protein or binding fragment thereof and a second binding protein or binding fragment thereof comprising the steps of: (a) transforming or transfecting a host cell with a first DNA expression construct comprising a first selectable marker gene encoding a first selectable marker protein and a repressor gene encoding a repressor protein, said repressor gene under transcriptional control of a promoter; (b) transforming or transfecting said host cell with a second DNA expression construct comprising a second selectable marker gene encoding a second selectable marker protein and a third selectable marker gene encoding a third selectable marker protein, said third selectable marker gene under transcriptional control of an operator, said operator specifically acted upon by said repressor protein such that interaction of said repressor protein with said operator decreases expression of said third selectable marker protein; (c) transforming or transfecting said host cell with a third DNA expression construct comprising a fourth selectable marker gene encoding a fourth selectable marker protein and a first fusion protein gene encoding a first binding protein or binding fragment thereof in frame with either a DNA binding domain of a transcriptional activation protein or a transactivating domain of said transcriptional activation protein; (d) transforming or transfecting said host cell with a fourth DNA expression construct comprising

- 9 -

a fifth selectable marker gene encoding a fifth selectable marker protein and a second fusion protein gene encoding a second binding protein or binding fragment thereof in frame with either the DNA binding domain of said transcriptional activation protein or the transactivating domain of said transcriptional activation protein, whichever is not included in first fusion protein gene; (e) growing said host cell under conditions which permit expression of said first binding protein or fragment thereof and said second binding protein or fragment thereof such that said first binding protein or fragment thereof and second binding protein or binding fragment thereof interact bringing into proximity said DNA binding domain and said transactivating domain reconstituting said transcriptional activating protein; said transcriptional activating protein acting on said promoter to increase expression of said repressor protein; said repressor protein interacting with said operator such that said third selectable marker protein is not expressed; (f) detecting absence of expression of said selectable gene; (g) growing said host cell in the presence of a test compound of binding between said first protein or fragment thereof and said second binding protein or fragment thereof; and (h) comparing expression of said selectable marker protein in the presence and absence of said test compound wherein decreased expression of said selectable marker protein is indicative of an ability of the test compound to inhibit binding between said first binding protein or binding fragment thereof and said second binding protein or binding fragment thereof such that said transcriptional activating protein is not reconstituted, expression of said repressor protein is not increased, and said operator increases expression of said selectable marker protein.

The methods of the invention encompass any and all of the variations in host cells as described above. In particular, the invention encompasses a method wherein: the host cell is a yeast cell; the selectable marker gene encodes HIS3; transcription of the selectable marker gene is regulated by the *tet* operator; the repressor protein gene encodes the

- 10 -

tetracycline resistance protein; transcription of the tetracycline resistance protein is regulated by the HIS3 promoter; the DNA binding domain is derived from LexA; and the transactivating domain is derived from VP16. In another embodiment, the invention encompasses a method wherein: the host cell is a yeast cell; the selectable marker gene encodes HIS3; transcription of the selectable marker gene is regulated by the *tet* operator; the repressor protein gene encodes the tetracycline resistance protein; transcription of the tetracycline resistance protein is regulated by the alcohol dehydrogenase promoter; the DNA binding domain is derived from LexA; and the transactivating domain is derived from VP16.

In alternative embodiments of the invention wherein the host cell is a mammalian cell, variations include the use of mammalian DNA expression constructs to encode the first and second recombinant fusion genes, the repressor gene, and the selectable marker gene, and use of selectable marker genes encoding antibiotic or drug resistance markers (i.e., neomycin, hygromycin, thymidine kinase).

There are at least three different types of libraries used for the identification of small molecule modulators. These include: (1) chemical libraries, (2) natural product libraries, and (3) combinatorial libraries comprised of random peptides, oligonucleotides or organic molecules.

Chemical libraries consist of structural analogs of known compounds or compounds that are identified as "hits" via natural product screening. Natural product libraries are collections of microorganisms, animals plants or marine organisms which are used to create mixtures for screening by: (1) fermentation and extraction of broths from soil, plant or marine microorganisms or (2) extraction of plants or marine organisms. Combinatorial libraries are composed of large numbers of peptides, oligonucleotides or organic compounds as a mixture. They are relatively easy to prepare by traditional automated synthesis methods, PCR, cloning or proprietary synthetic methods. Of particular interest are peptide and

- 11 -

oligonucleotide combinatorial libraries. Still other libraries of interest include peptide, protein, peptidomimetic, multiparallel synthetic collection, recombinatorial, polypeptide libraries.

The utility of the various aspects of the invention is manifest.

Host cells of the invention are useful to demonstrate *in vivo* binding capacity of both known and suspected binding partner proteins in a recombinant system. Such an expression system permits systematic analysis of the structure and function of a particular binding protein, thus permitting identification and/or synthesis of potential modulators of the physiological activity of the binding proteins. The methods of the invention are particularly useful to identify and improve molecules which are capable of inhibiting specific and general protein/protein interactions. Inhibitors identified by the methods of the invention can then be examined for utility *in vivo* as therapeutic and/or prophylactic medicaments for conditions associated with various protein/protein interactions.

Description of the Drawing

Figure 1 describes the mechanics of the split hybrid assays.

Detailed Description of the Invention

The present invention relates generally to methods designated split hybrid assays to identify inhibitors of protein/protein interactions and is illustrated by the following examples describing various methods for making and using the invention. In particular, Example 1 relates to construction of various plasmids and expression constructs utilized in the invention. Example 2 described generation of various yeast transformants used to identify inhibitor compounds. Examples 3, 4, 5 and 6 address use of the split hybrid assay to examine CREB/CBD binding, Tax/SRF binding, CKI/CREB binding and AKAP 79 binding to various partner protein, respectively. Example 7 describe general application of the split hybrid assay. Example 8 relates to

- 12 -

use of the split hybrid assay for weakly interacting binding partners. Example 9 describes general assay methods. Example 10 addresses use of the split hybrids assay to identify agents that prevent receptor desensitization and drug tachyphylaxis.

Example 1 Plasmid Construction

In the examples that follow, various plasmid constructs were utilized as described. To simplify discussion of the exemplified assays, this example describes construction of the various plasmids used in the following examples. For clarity, the plasmids are grouped according common features relating to their applications in the assays later discussed.

I. Plasmids Encoding Reporter Gene *HIS3*

A. pRS303/1xteop-MluI

One copy of the *ter* operator sequence was engineered into position -53 in the *HIS3* promoter of pRS313 [Sikorski, R.S. *et al.*, *Genetics* 122:19-27 (1989)] by using the polymerase chain reaction (PCR). Two primary PCR reactions using pRS313 as a template were performed which utilized a 5'-terminal oligonucleotide designated Eco47III-5' and a 3'-inner oligonucleotide designated Teop internal 3' to yield a primary 5'-PCR product and a 5'-inner oligonucleotide designated Teop internal 5' and a 3'-terminal oligonucleotide designated Nhe I 3' to yield a primary 3'-PCR product.

Eco47 III-5'	SEQ ID NO: 1
5'-TTGGTGAGCGCTAGAGATCAGTCAGCCAG	
Teop int. 3'	SEQ ID NO: 2
5'-TATACTCTATCAATGATAGAGTAATTCATATGATGATAAATGCC	
Teop int. 5'	SEQ ID NO: 3
5'-ATTACTCTATCATGATAGAGTATATAAAGTAATGATTTTC)	
Nhe I 3'	SEQ ID NO: 4

- 13 -

5'-AATTCTGCTAGCCTTGCAGCAAGC

5' and 3' inner oligonucleotides contain complementary sequence such that 3' sequence of the primary 5' PCR product overlaps with 5' sequence of the primary 3' PCR product. The 5' terminal oligonucleotide contains the restriction site *Eco47III* while the 3' terminal oligonucleotide contains the restriction site *NheI* in order to facilitate subsequent subcloning. The primary PCR reactions were performed with *Pfu* DNA polymerase (Stratagene, La Jolla, CA) using reaction conditions described by the manufacturer. PCR products were isolated by Bio101 (Visa, CA) Gene Clean III gel extraction. The primary 5' and 3' PCR products were then combined in a second PCR reaction and amplified using the 5' and 3' terminal oligonucleotides, *Eco47III*-5' and *NheI* 3'. The second PCR reaction was performed with *Veri* DNA polymerase (New England Biolabs, Beverly, MA) using reaction conditions described by the manufacturer, except that the reactions were supplemented with 4 mM Mg^{2+} . The final PCR product contained one *lei* operator sequence inserted into position -53 of the *HIS3* promoter and nucleotides 52-48 deleted in the construction. The final PCR product was isolated, digested with *Eco47III* and *NheI* and cloned into pRS313 previously digested with *Eco47III* and *NheI*. The resulting plasmid was designated pRS313/1xteop. DNA sequencing confirmed the presence of one copy of the *lei* operator sequence in pRS313/1xteop and confirmed integrity of the *Eco47III* and *NheI* junctions.

A *MluI* restriction enzyme site was engineered into position -22 in the *HIS3* promoter of pRS313/1xteop by utilizing PCR using *Veri* DNA polymerase using pRS313/1xteop as template. One PCR construct was amplified using the 5' terminal oligonucleotide *Eco47 III*-5' (SEQ ID NO: 1) containing an *Eco47III* restriction site and a 3'-oligonucleotide designated *Mlu I* 3' containing a *MluI* restriction site.

- 14 -

Mlu I 3' SEQ ID NO: 5
5'-CGCAGCGGTGGAAGAATTCACATTACTTATATA

A second PCR product was amplified using the 3'-terminal oligonucleotide *NheI* 3' (SEQ ID NO: 4) containing a *NheI* restriction site and a 5'-oligonucleotide designated *Mlu I* 5' containing a *MluI* restriction site.

Mlu I 5' SEQ ID NO: 6
5'-CGCACGCGTATATAAAAAATGAGCAGGCAAG

The first PCR product was isolated and digested with *Eco47III* and *MluI*, while the second PCR product was isolated and digested with *MluI* and *NheI*. These digested products were isolated and ligated in a triple ligation with pRS313 previously digested with *Eco47III* and *NheI*. The resulting plasmid was designated pRS313/1xteop-*MluI*. DNA sequencing confirmed the presence of the *MluI* site in pRS313/1xteop-*MluI* and confirmed that integrity of the *Eco47III* and *NheI* junctions were maintained.

A pRS303/1xteop-*MluI* plasmid was constructed by first removing the *Eco47III*/*NheI* fragment containing the altered *HIS3* promoter from the pRS313/1xteop-*MluI* vector and ligating the isolated fragment into pRS303 previously digested with *Eco47III* and *NheI*. DNA sequencing confirmed proper insertion of the *Eco47III*/*NheI* fragment.

20 B. pRS303/2xteop-LX52

One copy each of the *lei* operator sequence was engineered into positions -53 and -22 in the *HIS3* promoter of pRS303 (Sikorski, *et al.*, *Genetics* 122:19-27 (1989)). PCR was utilized to engineer one copy into position -53 which resulted in plasmid pRS303/1xteop. To insert the second copy, a *MluI* site was introduced at position -22 in the *HIS3* promoter using PCR. The new plasmid was designated pRS303/1xteop-*MluI*.

The *lei* operator was created by annealing two complementary

- 15 -

oligonucleotides tetop-1 and tetop-2.

SEQ ID NO: 7
tetop-1
5'-CGCGTACTCTATCATTTGATAGAGTA;
SEQ ID NO: 8
tetop-2
5'-ATGAGATAGTAAGTATCTCATTCGCGC

When annealed, the *tet* operator sequence contains flanking *Mlu*I sites. Both oligonucleotides were phosphorylated using T4 polynucleotide kinase (Gibco BRL, Grand Island, NY) at 37°C for one hour and annealed by first heating at 70°C for 10 minutes and then cooling to room temperature. The annealed oligonucleotides were isolated and ligated into pRS303/1xtetop-*Mlu*I previously digested with *Mlu*I. The resulting plasmid was designated pRS303/2xtetop. DNA sequencing confirmed insertion of one copy of the *tet* operator sequence in the *Mlu*I site.

The *LYS2* gene was digested from pLYS2 [Hollenberg, S.M. *et al.*, *Mol. Cell. Biol.* 15:3813-3822 (1995)] with *Eco*RI and *Hind*III and the isolated fragment blunt ended using the large fragment of DNA polymerase I (Gibco BRL, Grand Island, NY). Phosphorylated *Sst*I linkers (New England Biolabs, Beverly, MA) were ligated to the fragment, the fragment digested with *Sst*I, and the resulting fragment ligated into pRS313 previously digested with *Sst*I. The resulting plasmid was designated pRS313/LYS2.

The *LYS2* fragment was removed from pRS313/LYS2 with *Sst*I digestion and inserted into pRS303/2xtetop previously digested with *Sst*I. The resulting plasmid was designated pRS303/2xtetop-LYS2.

Similarly, the *LYS2* *Sst*I fragment was inserted into pRS303/1xtetop-*Mlu*I previously digested with *Sst*I yield pRS303/1xtetop-*Mlu*I-LYS2.

C. pRS303/3xtetop-LYS2

- 16 -

Two copies of the *tet* operator sequence were created by self-annealing a palindromic oligonucleotide Tetop 2x with itself.

SEQ ID NO: 9
Tetop 2x
5'-CGCGTACTCTATCATTTGATAGAGTCTAGACTTATCATATGATAGAGTA

The annealed oligonucleotide contained flanking *Mlu*I sites. The oligonucleotide was phosphorylated, annealed, and isolated as above. The isolated annealed and *Mlu*I-digested oligonucleotide was ligated into pRS303/1xtetop-*Mlu*I-LYS2 previously digested with *Mlu*I to yield pRS303/3xtetop-LYS2. The presence of two copies of the *tet* operator sequence in the *Mlu*I site was confirmed by DNA sequencing.

D. pRS303/4xtetop-LYS2 and pRS303/8xtetop-LYS2

Three or seven copies of the *tet* operator were created using PCR with *Ver*u DNA polymerase as described above. Plasmid pUHC-13-3 [Grossen and Bujang, *Proc. Natl. Acad. Sci. (USA)* 89:5547-5551 (1992)] was used as template DNA using 5'- and 3'- oligonucleotides, *Mlu*I/Sph I 5' and *Mlu*I Sph I 3', containing an exterior *Mlu*I restriction enzyme site nested internally by a *Sph*I restriction enzyme site.

SEQ ID NO: 10
*Mlu*I/Sph I 5'
5'-GCGACGCGTCATGCGCGTCTTCAAGAATTCCCTCGAG
SEQ ID NO: 11
*Mlu*I Sph I 3'
5'-GCGACGCGTGCATGCCACCGTACACGCGCTACTCGA

The PCR products were separated on an agarose gel and the ladder of different sized DNA fragments was isolated, digested with *Mlu*I, and ligated into the *Mlu*I restriction site of pRS303/1xtetop-*Mlu*I-LYS2. DNA sequencing revealed that either three or seven copies of *tet* operators were inserted into the *Mlu*I site of pRS303/1xtetop-*Mlu*I-LYS2 to provide either pRS303/4xtetop-LYS2 or pRS303/8xtetop-LYS2.

- 17 -

E. PRS303/6x1etop-LYS2 and PRS303/10x1etop-LYS2

A *SphI* restriction enzyme site was introduced at position -85 in the *HIS3* promoter of pRS303/3xetop-LYS2 using PCR with *VenI* DNA polymerase as described. Plasmid pRS303/3xetop-LYS2 was used as a template DNA. A first fragment was amplified using the 5'-terminal oligonucleotide Ecoa47 III-5' (SEQ ID NO: 1) described above containing an *Ecoa47III* restriction site and a 3'-oligonucleotide *Sph* I 3' containing a *SphI* restriction site.

SEQ ID NO: 12	Sph I 3'
5'-CATGGCATGCAAAAAAAGAGTCATCCGCTAGG	10

A second PCR product was amplified using the 3'-terminal oligonucleotide N1e 1 3' (SEQ ID NO: 4) described above containing a *MneI* restriction site and a 5'-oligonucleotide containing a *SphI* restriction site.

Sph I 5' SEQ ID NO: 13
5'CATGGCATGCTTAGCGATTGGCATTATCACAT

The PCR products were isolated as described above. The first PCR product was digested with *Eco47III* and *SphI*, and the second PCR product was digested with *SphI* and *MfeI*. Both digestion products were ligated in a triple ligation along with pRS303/3xetop-LYS2 previously digested with both *Eco47III* and *MfeI*. The resulting plasmid was designated pRS303/3xetop-SphI-LYS2. The presence of the *SphI* site in pRS303/3xetop-SphI-LYS2 was confirmed by DNA sequencing analysis.

Three copies of *ter* operators were isolated as a single fragment by digesting pRS303/4x_{topo}-LYS2 with *Sph*I. The isolated fragment was ligated into the *Sph*I site of pRS303/3x_{topo}-*Sph*I-LYS2 to yield pRS303/6x_{topo}-LYS2. The presence of three additional copies of the *ter* operator in pRS303/6x_{topo}-LYS2 at the *Sph*I site was confirmed by DNA

- 18 -

F. PRS313/MluI and PRS303/MluI

Seven copies of *tet* operators were isolated as a single fragment by digesting pRS303/8xetop-LYS2 with *Sph*I. The isolated fragment was ligated into the *Sph*I site of pRS303/3xetop-*Sph*-LYS2 to yield pRS303/10xetop-LYS2. The presence of seven additional copies of the *tet* operator in pRS303/10xetop-LYS2 at the *Sph*I site was confirmed by DNA sequencing.

A *Mlu*I restriction enzyme site was engineered into position -22 in the *HIS3* promoter of pRS313 utilizing PCR and *Ven*I DNA polymerase as

15 a *MluI* restriction site. A second PCR product was amplified using the 3' terminal oligonucleotide Nhe I 3' (SEQ ID NO: 4) containing a *MneI* restriction site and the 5' oligonucleotide Mlu I 5' (SEQ ID NO: 6) containing a *MluI* restriction site. The first PCR product was isolated and digested with *Eco4VII* and *MluI*, while the second PCR product was isolated and digested with *MluI* and *MneI*. The digested products were partially purified and joined in a triple ligation with pRS313 which had been previously digested with *Eco4VII* and *MneI*. The resulting plasmid was designated pRS313/*MluI*. DNA sequencing confirmed the presence of the *MluI* site in pRS313/*MluI* and to confirm the integrity of the *Eco4VII* and *MneI* junctions.

25 pRS303/MluI was constructed in exactly the same manner as
pRS313/MluI except that pRS303 was used in place of pRS313.

G. PR\$313/1xtetop

See above wherein PRS313/1xtelop is an intermediate in the

- 19 -

construction of pRS303/1xteop-MluI.

H. pRS313/MluI-1xteop and pRS303/MluI-1xteop

One copy of the *ter* operator sequence was created by annealing two complementary oligonucleotides tetop-1 and tetop-2 (SEQ ID NO: 7 and SEQ ID NO: 8). The annealed *ter* operator sequence contains flanking *MluI* sites. The oligonucleotides were phosphorylated using T4 polynucleotide kinase (Gibco BRL, Grand Island, NY) at 37°C for one hour and annealed by first heating at 70°C for 10 minutes followed by cooling to room temperature. The annealed oligonucleotides were isolated and ligated separately into *MluI*-digested pRS313/MluI and pRS303/MluI, the resulting plasmids being designated pRS313/MluI-1xteop and pRS303/MluI-1xteop. DNA sequencing confirmed the presence of one copy of the *ter* operator in the *MluI* sites of both plasmids.

In order to produce plasmids bearing multiple copies of the *ter* operator, annealed oligonucleotides described above were ligated together overnight at 16°C. After isolation of the ligation products, they were inserted into the *MluI* of pRS313/MluI. DNA sequencing analysis confirmed that one clone, pRS313/MluI-4xteop, was produced which contained four copies of *ter* operator in the *MluI* site. However, upon further examination of this clone it was discovered that it had been subjected to a recombination event and was therefore not useful for further cloning steps. Continued attempts to insert multiple copies of the *ter* operator into the *MluI* site of pRS313/MluI by ligating multimers of the *ter* operator have been unsuccessful.

I. pRS313/1xteop-MluI

See above wherein construction of pRS313/1xteop-MluI was an intermediate in the construction of pRS303/1xteop-MluI.

- 20 -

J. pRS313/2xteop

One copy of the *ter* operator sequence was created using annealed complementary oligonucleotides tetop-1 and tetop-2 (SEQ ID NO: 7 and SEQ ID NO: 8). Annealed oligonucleotides were ligated into the *MluI* site of pRS313/1xteop-MluI to yield pRS313/2xteop. DNA sequencing confirmed the presence of two copies of the *ter* operator in the *MluI* site.

K. pRS303/2xteop

See above wherein pRS303/2xteop was an intermediate in the construction of pRS303/2xteop-LYS2.

L. pRS313/LYS2 and pRS313/LYS2

The *LYS2* gene was digested from pLYS2 with *EcoRI* and *HindIII* digestion. The *EcoRI/HindIII* fragment was blunt ended using the large fragment of DNA polymerase I (Gibco BRL, Grand Island, NY) and ligated with phosphorylated *SrfI* linkers (New England Biolabs, Beverly, MA). The resulting fragment was digested with *SrfI* and ligated into pRS313 previously digested with *SrfI*. The resulting plasmid was designated pRS313/LYS2. Because the *LYS2* fragment was shown to have inserted into pRS313 in both orientations, plasmids with the *LYS2* gene in both orientations were transformed separately into the yeast strain SEY6210c (*MAT α _leu2-3,112 ura3-52 his3- Δ 200 trp1- Δ 501 lys2-801 ura2- Δ 9* [Robinson *et al.*, *Mol. Cell. Biol.* 8:4936-4948 (1988)]). Both clones allowed the yeast to grow in the absence of lysine indicating that orientation of the *LYS2* gene in pRS313 did not affect the expression of an active gene.

The *LYS2* fragment was removed from pRS313/LYS2 with *SrfI* and ligated into the *SrfI* site of:

pRS313/1xteop-MluI giving plasmid pRS313/1xteop-MluI-LYS2,
pRS313/2xteop giving plasmid pRS313/2xteop-LYS2.

- 21 -

PRS303/1xiciop-MluI giving plasmid PRS303/1xiciop-MluI-LYS2, and PRS303/2xiciop giving plasmid PRS303/2xiciop-LYS2.

II. Plasmids Encoding Reporter Gene *TeR*

A. PRS306/HIS3:*TeR*/Term

The 5' promoter sequence of the yeast *HIS3* gene, encompassing nucleotides -75 to +23, was ligated to the translational start of *TeR*. In addition, the DNA sequence encoding the simian virus 40 (SV40) large T antigen nuclear localization signal was ligated in frame with the nucleotide sequence encoding the last amino acid residue of *TeR*. The chimeric fragment was created by the same PCR strategy as described above.

The *HIS3* promoter fragment, the primary 5'-PCR product, was amplified by PCR from plasmid p601 [Gruenberg, D.A., *Science* 257:1089-1095 (1992)] using a 5'-terminal oligonucleotide T7 Promoter primer and a 3'-inner oligonucleotide 3'-*TeR* inner primer.

15 T7 Promoter primer SEQ ID NO: 14
5'-TAATACGACTCACTATATAGGG

3'-*TeR* inner primer SEQ ID NO: 15
5'-TCTAGACTTGCCCTCGTTATC

20 The primary 3' PCR product containing the *TeR* coding sequence was amplified from pSLF104 [Forsburg, *Nucl. Acid. Res.* 21:2955-2956 (1993)] with a 5'-inner oligonucleotide 5'-*TeR* inner primer and a 3'-terminal oligonucleotide 3'-*TeR* terminal primer.

5'-*TeR* inner primer SEQ ID NO: 16
5'-CGAAGGCAAGAATGCTAGATTAGATAAAG

25 3'-*TeR* terminal primer SEQ ID NO: 17
5'-CGCGGATCGCTTCTCTTTTGGAGACCCACTTCACATTAAAG

- 22 -

An *EcoRI* site derived from the p601 fragment and a *BamHI* site in the 3'-terminal oligonucleotide were used in subsequent subcloning. The PCR products were gel-purified and amplified in a second PCR reaction with 5'- and 3'-terminal oligonucleotides, T7 Promoter primer (SEQ ID NO: 14) and 3'-*TeR* terminal primer (SEQ ID NO: 17). The secondary PCR product was isolated, digested with *EcoRI* and *BamHI*, and ligated into PRS306/Term previously digested with *EcoRI* and *BamHI*. The resulting plasmid was designated PRS306/HIS3:*TeR*/Term which comprises the complete *TeR* coding sequence in frame with sequences encoding the nuclear localization signal of SV40 large T antigen.

B. PRS316/HIS3:*TeR*/Term

The construction protocol for this plasmid was the same as described above for subcloning a *HIS3* DNA into PRS306/Term except that the vector for subcloning was PRS316/Term described above.

15 C. PRS306/1xLexAop/HIS3:*TeR*

Oligonucleotides LexAop (100a) and LexAop (100b) containing a single copy of LexA operator were phosphorylated with 74 polynucleotide kinase (Gibco BRL, Grand Island, NY) at 37°C for one hour.

20 LexAop (100a) SEQ ID NO: 18
5'-AATTGCTCGAGTACTGTATGTACATACAGTAG

LexAop (100b) SEQ ID NO: 19
5'-AATTCTACTGTATGTACATACAGTACTGAGC

25 Following phosphorylation, the oligonucleotides were annealed by heating at 70°C for 10 minutes followed by cooling to room temperature. The annealed oligonucleotide containing 5' and 3' *EcoRI* overhanging ends was subcloned into PRS306/HIS3:*TeR*/Term previously digested with *EcoRI*. The number

- 23 -

of copies of inserted oligonucleotide was confirmed by DNA sequencing. The plasmid containing a single copy of the LexA operator was designated PRS306/1xLexAop/HIS3::TeR.

D. PRS316/2xLexAop/HIS3::TeR

5 The subcloning protocol for this construct was the same as described above for PRS306/1xLexAop/HIS3::TeR. The annealed oligonucleotides encoding the LexA operator included overhanging *EcoRI* ends and during ligation, the individual annealed fragments were able to multimerize, inserting into the parental plasmid more than one copy of the desired LexA sequence. The number of copies of inserted oligonucleotides was confirmed by DNA sequencing.

E. PRS306/2xLexAop/HIS3::TeR

15 A DNA fragment containing two copies of LexA operator and the chimeric *HIS3::TeR* reporter was excised from PRS316/2xLexAop/HIS3::TeR by digestion with *KpnI* and *BamHI* restriction enzymes. The fragment was gel-purified and subcloned into PRS306/Term previously digested with *KpnI* and *BamHI* and the resulting construct was sequenced to confirm the presence of two copies of the LexA operator.

F. PRS306/4xLexAop/HIS3::TeR and PRS306/8xLexAop/HIS3::TeR

20 A pair of oligonucleotides SH101A and SH101B were utilized in PCR to amplify the LexA binding site multimer from the plasmid SH18-34ASpe [Hollenberg, S.M., *et al.*, *Mol. Cell Biol.* 15:3813-3822 (1995)].

25 SH101A
5'-CCGGAATTCTCGAGACATATCCATATCTAATC
SH101B
5'-CCGGAATTCACATAATCGCATATATCATC
SEQ ID NO: 21

- 24 -

The amplification product containing four copies of LexA operator was gel-purified, digested with *EcoRI*, and subcloned into PRS306/HIS3::TeR/Term previously digested with *EcoRI*. The number of LexA operators were determined by DNA sequencing.

G. PRS306/8xLexAop/HIS3::TeR

5 A PCR strategy was used to link the 5' promoter sequence of the yeast *HIS3* gene encompassing nucleotides -75 to +23 to the translational start of *TeR*. Sequences encoding the SV40 large T antigen nuclear localization signal were fused in frame with the nucleotide sequence encoding the last amino acid residue of *TeR*. The PCR product was digested with *EcoRI* and *BamHI* and inserted into PRS306/Term previously digested with *EcoRI* and *BamHI*. The resulting plasmid was designated PRS306/HIS3::TeR/Term, and was shown to encode the complete *TeR* protein in frame with the nuclear localization signal of SV40 large T antigen. The fusion protein is followed by four amino acids generated by the vector backbone (Arg-Ile-His-Asp).

15 The LexA binding site multimer from the plasmid pSH18-34ASpe [Hollenberg, S.M. *et al.*, *Mol. Cell Biol.* 15:3813-3822 (1995)] was amplified by PCR, digested with *EcoRI*, and subcloned into the *EcoRI* site of PRS306/HIS3::TeR/Term resulting in plasmid PRS306/8xLexAop/TeR.

H. PADH/TeR

20 The DNA coding sequence of *TeR* was amplified by PCR from pSLF104 using two oligonucleotides, NcoI-*TeR* and 3'-*TeR* terminal primer (SEQ ID NO: 17).

25 NcoI-*TeR*
5'-CATGCCATGGCCATGTCTAGATTGATATAAAG
SEQ ID NO: 22

- 25 -

The resulting product was gel-purified, digested with *NcoI* and *BamHI*, and subcloned into a pBTM116 [Bartel, *et al.*, in Cellular Interactions in Development: a Practical Approach, Hartley (ed.), IRL Press, Oxford, pp. 153-179 (1993)] shuttle vector containing an ADH promoter, previously digested with *NcoI* and *BamHI*. For construction of this vector, DNA generated by PCR and DNA obtained by restriction enzyme digestion of the polylinker region in plasmid pBluescript (Stratagene, La Jolla, California) were used to engineer additional restriction sites 5' and 3' of the ADH promoter. The *TeiR* protein encoded from this construct is expressed containing additional amino acids Met⁻²-Ala⁻¹ before the initiating methionine and also contains the nuclear localization signal of SV40 large T antigen located after the last amino acid of *TeiR* as described above.

I. PRS306/ADH:TeiR/Term

A fragment encoding the ADH promoter and *TeiR* was removed from plasmid pADH:TeiR with *XhoI* and blunted-ended with the large fragment of DNA polymerase I (Gibco BRL, Grand Island, NY). *EcoRI* linkers (New England Biolabs, Beverly, MA) were added and the fragment was digested with *EcoRI* and *BamHI*. The resulting fragment was gel-purified and ligated into PRS306/Term previously digested with *EcoRI* and *BamHI*.

J. PRS306/4xLexAop/ADH::TeiR and PRS306/8xLexAop/ADH::TeiR

The subcloning protocol used to insert multiple copies of the LexA operator into PRS306/ADH:TeiR/Term was the same as described previously for PRS306/4xLexAop/HIS3:TeiR and PRS306/8xLexAop/HIS3:TeiR.

- 26 -

III. Plasmids Encoding Binding Proteins

A. pLexA-CBD

A DNA fragment containing the CREB binding domain of CBP (CBD), amino acids 461-682, was PCR amplified from plasmid CBP-0.8 [Chrivia, J.C. *et al.*, *Nature* 365:835-839 (1993)] using a pair of oligonucleotides designated 5' CBD primer and 3' CBD primer.

5' CBD primer

SEQ ID NO: 23

5'-GCGAATTCCGCCAGGCGACACAGAATCCACT

3' CBD primer

SEQ ID NO: 24

5'-CGGATCTCTGGCTGTTACCCAGGATGCCTTG

Following gel purification, the amplification product was digested with *EcoRI* and *BamHI*, and ligated into plasmid pBTM116 [Bartel, *et al.*, in *Cellular Interactions in Development: a Practical Approach*, (ed) Hartley, D.A. (IRL Press, Oxford), pp. 153-179 (1993)] previously digested with *EcoRI* and *BamHI*.

B. pVP16-CBD

A DNA fragment encoding the CBP sequence was excised from pLexA-CBD by digestion with *EcoRI* and *BamHI*. Plasmid pLexA-CBD was linearized with *EcoRI* digestion, the resulting overhanging ends blunt-ended using the Klenow fragment of DNA polymerase I, and the ends ligated with *BamHI* linkers. The resulting fragment was inserted into pVP16 [Hollenberg, *et al.*, *Mol. Cell. Biol.* 15:3813-3822 (1995)] previously digested with into *BamHI*.

- 27 -

C. **pVP16-CREB**

Plasmid pCDNA3/CREB283 [Sun and Maurer, *J. Biol. Chem.* 270:7041-7044 (1995)], containing the VP16 transactivation domain fused to sequences of the rat CREB transactivation domain (1 to 283 aa) was linearized with *XhoI* and *BamHI* linkers (New England Biolabs) ligated to the resulting blunt-ended *XhoI* sites. DNA encoding the VP16/CREB chimeric protein was removed with *HindIII* and *BamHI* digestion and following gel purification, ligated into the *HindIII* and *BamHI* sites of pVP16 which encodes the *LEU2* gene.

D. **pVP16-CREB(BglII-SacII)-LacZ**

A DNA fragment encoding β -galactosidase was PCR amplified from plasmid pSV- β -galactosidase vector (Promega, Madison, WI) using a pair of oligonucleotides, 5' β -gal primer and 3' β -gal primer and inserted into the *NcoI* site of pVP16 to produce pVP16-LacZ.

5' β -gal primer

5'-ATGTTACCGCGCGCTAGTCGTTTACACAGTCGTGAC

3' β -gal primer

5'-ATGTTACCGCGCGCTATTTTTGACACAGACCAAC

SEQ ID NO: 29

SEQ ID NO: 30

20

A PCR fragment containing CREB sequences encoding amino acid residues 1 to 283 was amplified from plasmid pRSV-CREB341 [Kwok, *et al.*, *Nature* 380: 642-646 (1996)] using a pair of oligonucleotides, 5' CREB 341 primer and 3' CREB 283 primer, and inserted into pVP16-LacZ vector at the *BamHI* site.

25

5' CREB 341 primer

5'-CGCGGATCCGGATGACCATGACTCTGGAG

3' CREB 283 primer

5'-CGCGGATCCGCTGCTCTTTCAGCAGGCTG

SEQ ID NO: 25

SEQ ID NO: 28

- 28 -

To generate a cassette vector for producing and subcloning mutated CREB sequences as described below, PCR was used to engineer a *BglII* site using oligonucleotides 5' *BglII* primer and 3' *BglII* primer, at nucleotides 273 to 278 and a *SacII* site using oligonucleotides 5' *SacII* primer and 3' *SacII* primer at nucleotides 500 to 505 of the CREB activation domain.

5' *BglII* primer

5'-CGGAGATCTAAGAGACTTTCTCCGGAAGCTCAG

SEQ ID NO: 31

3' *BglII* primer

5'-CGGAGATCTTACAGGAAGACTGAAGCTGT

SEQ ID NO: 32

5' *SacII* primer

5'-CCACCGCGGACAGTGCCAAACCCGATTAC

SEQ ID NO: 33

3' *SacII* primer

3'-CATCCGCGGTGTGATGCGCAGGGGCTGA

SEQ ID NO: 34

E. **pLacZ-CREB 283**

A DNA fragment containing the rat CREB transactivation domain (amino acids 1 to 283) was excised from pCDNA/CREB283 [Sun and Maurer, *supra*] with *SmaI* and *XbaI* digestion. The 5' *XbaI* site was blunt ended with the large fragment of DNA polymerase I (Gibco BRL, Grand Island, NY) and *SacI* linkers (New England Biolabs, Beverly, MA) added. The fragment was digested with *SacI* and subcloned into the *SacI* site of pBTM116.

F. **pLacZ-CREB 341**

A DNA fragment containing the rat CREB 341 cDNA was amplified by PCR from pCDNA/CREB341 [Kwok, *supra*] using a pair of oligonucleotides, 5' CREB 341 primer (SEQ ID NO: 25) and 3' CREB 341 primer.

- 29 -

3' CREB 341 primer SEQ ID NO: 26
5'-CGCGGATCCTTAATCTGACTTGTCGACAGTA

After gel purification, the PCR product was digested with *Bam*HI, and subcloned into the *Bam*HI site of pBTM116.

5 G. pLexA-CREB₃₄₁-M1

A DNA fragment containing the rat CREB sequence with a mutation changing serine at position 133 to alanine was amplified by PCR from plasmid Rc/RSV CREB-M1 [Kwok, *et al.*, *supra*] using the same set of primers as described for pLexA-CREB 341. 5' CREB 341 primer (SEQ ID NO: 25) and 3' CREB 341 primer (SEQ ID NO: 26). The resulting amplification product was gel-purified, digested with *Bam*HI, and subcloned into the *Bam*HI site of pBTM116.

H. pVP16-CREB-M1

A PCR fragment containing CREB sequences coding for amino acid residues 1 to 283 including the serine 133 mutation to alanine was amplified using a pair of oligonucleotides, 5' CREB 283 primer and 3' CREB 283 primer (SEQ ID NO: 28). The PCR fragment was gel-purified, digested with *Bam*HI and inserted into the *Bam*HI site of pVP16.

5' CREB 283 primer SEQ ID NO: 27
5'-CGCGGATCCTCCATGACCATGGAATCTGGAGCC

I. pLexA-SRF

A DNA fragment containing human SRF was excised from plasmid pCGN-SRF [Gruenberg, D.A., *et al.*, *Science*, 257:1089-1095 (1992)] with *Xho*I and *Bam*HI digestion. The *Xho*I site of the fragment was blunt-ended by the large fragment of DNA polymerase I (Gibco BRL, Grand Island, NY), ligated with *Bam*HI linkers, digested with *Bam*HI, and inserted

- 30 -

into pBTM116 previously digested with *Bam*HI.

J. pVP16-Tax

A DNA sequence encoding full length Tax protein was excised from pS6424 [Kwok, R.P.S., *et al.*, *Nature* 380:642-646 (1996)] with *Bam*HI digestion and was inserted into pVP16 previously digested with *Bam*HI.

IV. Plasmids For Binding Protein Controls

A. pLex

Plasmid pVP16 was digested with *Hind*III and *Bam*HI to remove the fragment encoding the VP16 transactivation domain. The digested vector was blunt-ended and self-ligated.

B. pLexA-VP16

The VP16 transactivation domain was PCR amplified from pGal-VP16 [Sadowski, *et al.*, *Nature* 335:563-564 (1988)] with a pair of oligonucleotides, 5'-VP16SH and 3'VP16SH and the resulting amplification product was digested with *Cla*I, blunt-ended, and inserted into pBTM116.

5'-VP16SH SEQ ID NO: 35
GGCTATCGATACGGCCCCCGGACCGAT
3'-VP16SH SEQ ID NO: 36
GGCTATCGATCTACCCACCGTACTTCGTC

C. pLexA-Lamin

See Hollenberg, S.M. *et al.*, *Mol. Cell. Biol.* 15:3813-3822 (1995).

- 31 -

V. Plasmids Encoding Reporter Gene Controls

A. pRS306/Term

The alcohol dehydrogenase (ADH) terminator sequence was excised from plasmid pBTM116 [Barrel, *et al.*, in *Cellular Interactions in Development: a Practical Approach*, (ed) Hartley, D.A. (IRL Press, Oxford), pp. 153-179 (1993)] with *SphI* and *PstI* restriction enzymes and both 3'-overhanging sequences were blunted by 74 DNA polymerase (Gibco BLR, Grand Island, NY). The fragment was gel-purified and subcloned into the blunt-ended *NotI* site in pRS306 [Sikorski and Hieter, *Genetics* 122:19-27 (1989)]. The orientation of inserted fragment was determined by DNA sequencing.

B. pRS316/Term

The subcloning protocol for inserting the ADH terminator sequence into pRS316 was the same as described for inserting the ADH sequence in pRS306.

Example 2

Generation of Yeast Assay Transformant

Selection of an appropriate yeast assay strain is an empirical determination based on growth characteristics of the transformed alternatives.

A general method to make the appropriate selection is described as follows.

Candidate yeast assay strains were transformed individually with reporter gene constructs and/or a plasmid encoding one of the experimental binding proteins. Assay strains thus transformed were then compared for relative differences in growth characteristics, with an optimal assay strain showing negligible growth on media lacking histidine and vigorous growth on media containing histidine. In practical application of this first step in selection using various plasmids transformed into assay strain Y1584, the following results were observed.

- 32 -

When the plasmid pLexA-VP16 encoding both the LexA DNA binding domain and the VP16 transactivating domain as a single protein was introduced into the assay cells, growth in the absence of histidine in the media was significantly reduced three days after transformation.

In assays including transformation with plasmids encoding multiple copies of the *tet* operator upstream of the *HIS3* gene, the following plasmids were separately utilized:

pRS303/1x*teop-HIS* (encoding a single *tet* operator sequence),
 pRS303/2x*teop-HIS* (encoding two *tet* operator sequences),
 pRS303/3x*teop-HIS* (encoding three *tet* operator sequences),
 pRS303/4x*teop-HIS* (encoding four *tet* operator sequences),
 pRS303/6x*teop-HIS* (encoding six *tet* operator sequences),
 pRS303/8x*teop-HIS* (encoding eight *tet* operator sequences), or
 pRS303/10x*teop-HIS* (encoding ten *tet* operator sequences).

In the assay strains transformed with plasmids encoding either one, two, or three copies of the *tet* operator upstream from the *HIS3* gene, cells grew on media lacking histidine at a rate similar to cells grown on media containing histidine. In yeast assay strains transformed with plasmids encoding either six, eight, or ten copies of the *tet* operator upstream from the *HIS3* gene, cell growth was low suggesting that these strains would not be useful in assays to examine binding and interruption of binding between test proteins. These results suggested that, in assay strains transformed with a reporter plasmid having more than three *tet* operator sequences upstream from the *HIS3* gene, normal activity of the *HIS3* promoter is disrupted and that these plasmids would not be useful.

In assays wherein yeast cells were transformed with only reporter plasmids (and not plasmids encoding binding partner fusion proteins) encoding multiple copies of the LexA operator 5' of the *TetR* gene, the

- 33 -

following results were observed. Growth of assay cells transformed with plasmids bearing one, two, four, and eight copies of the regulatory LexA operator upstream of the TetR gene appeared to be "copy number" dependent. Yeast cells transformed with plasmids having two copies of the LexA operator grew at a rate significantly higher than those assay cell transformed with a plasmid bearing only one copy of the operator. Cells transformed with plasmids encoding either four or eight LexA operators upstream of the TetR gene grew at an approximately equal rate, and better than assay cells bearing a TetR gene driven by two copies of the operator.

When the alcohol dehydrogenase (ADH) promoter was included upstream of the LexA operator (plasmids encoding either four or eight LexA operators) in the various reporter gene constructs, cell viability was the lowest.

The various cell lines constructed by the methods described above are shown in Table I, wherein various transformed yeast strains are identified (Strain #) along with the number of LexA operator sequences in the plasmid encoding TetR, the number of tetracycline operator sequences regulating expression of HIS3, and relative growth rate of the transformed strain on media containing histidine. It is important to note that growth variation of transformed cells in media containing histidine is observed, even in cell lines identically transformed. The number of "+" signs in Table I is indicative of the host cell's relative ability to grow on media lacking histidine in the absence of transformation with plasmids encoding potential binding proteins. Also in Table I, a subscript "a" is indicative of transformation with a plasmid bearing the alcohol dehydrogenase promoter; absence of a subscript "a" indicates use of the HIS3 promoter.

- 34 -

Table I
Various Yeast Transformants

5	Diphida L40 Strain # LexA TetOp His+	Y1002 4X 6X Y1079 1X 2X Y1570 1X 2X Y1580 2X 2X Y1582 2X 2X	+++ +++ +++ +++ +++
10	Diphida L40 Strain # LexA TetOp His+	Y1583 4X 2X Y1585 4X 2X Y1587 4X 2X Y1589 4X 2X	+++ +++ +++ +++
15	Y1584 8X 2X Y1586 8X 2X Y1588 8X 2X Y1590 8X 2X	+++ +++ +++ +++	
20	Diphida L40 Strain # LexA TetOp His+	Y1591 2X 2X Y1593 2X 2X Y1597 2X 4X Y1613 2X 4X	+++ +++ +++ +++
25	Y1616 2X 4X Y1600 2X 6X Y1606 2X 6X Y1627 2X 6X	+++ ++ + +++	
30	Y1603 2X 10X Y1621 2X 10X Y1609 2X 10X Y1624 2X 10X	+ ++ ++ ++	
35	Y1593 4X 2X Y1595 4X 2X Y1599 4X 4X Y1634 4X 4X Y1638 4X 4X	+++ +++ - + +	
Diphida L40 Strain # LexA TetOp His+	Y1604 4X 6X Y1607 4X 6X Y1628 4X 6X Y1632 4X 6X	+++ +++ +++ +++	
Diphida L40 Strain # LexA TetOp His+	Y1605 4X 10X Y1610 4X 10X Y1622 4X 10X Y1626 4X 10X	.. ++ ++ ++	
Diphida L40 Strain # LexA TetOp His+	Y1592 8X 2X Y1596 8X 2X Y1598 8X 2X Y1601 8X 4X Y1603 8X 4X Y1604 8X 6X Y1629 8X 6X Y1631 8X 6X	+++ +++ +++ ++ ++ ++ ++ ++	
Diphida L40 Strain # LexA TetOp His+	Y1614 8X 10X Y1611 8X 10X Y1623 8X 10X Y1625 8X 10X	+ ++ ++ ++	
Diphida L40 Strain # LexA TetOp His+	Y1605 8X 3X Y1607 8X 3X Y1611 8X 3X Y1609 8X 3X Y1624 8X 3X Y1601 8X 2X Y1614 8X 6X	+++ +++ +++ +++ +++ +++ +++	
Diphida L40 Strain # LexA TetOp His+	Y1604 4X 3X Y1606 4X 3X Y1607 4X 3X Y1608 4X 3X Y1625 4X 3X Y1604 8X 10X Y1611 8X 10X Y1623 8X 10X	+++ +++ +++ +++ +++ +++ +++ +++	

- 35 -

Example 3 CREB/CBP Binding Interaction

Use of the split-hybrid assay for studies of protein/protein binding wherein one of the binding components is randomly mutagenized was carried out using CREB and CBP binding proteins. The binding of CREB to CBP has been shown to require the phosphorylation of the CREB serine residue at position 133 in a region designated the "kinase-inducible domain" (KID) [Chirivia, *et al.*, *Nature* 365, 855-859 (1993); Kwok, *et al.*, *Nature* 370, 223-226 (1994)]. Functionally, changing serine at position 133 to alanine (a mutant designated CREB-M1) abolishes the ability of CBP to activate CREB-mediated transcription. Preliminary studies have indicated that the CREB-M1 mutant in the split-hybrid system prevents the interaction with CBP and subsequent growth of the yeast assay strain on media lacking histidine. Precisely what other portions of the KID of CREB are required for binding to CBP is unknown, however. To define other potentially important amino acid residues, the KID (amino acid residues 102 to 160) of CREB 341 was randomly mutagenized using PCR.

A. PCR Mutagenesis and Creation of Mutant Library

The technique used for mutagenic PCR was a modification of that described by Uppaluri and Towle [*Mol. Cell. Biol.* 15, 1499-1512 (1995)]. The reaction mixture contained 20 ng of pVP16-CREB(*Bgl*II-*Sac*II)-LacZ, 16 mM (NH₄)₂SO₄, 67 mM Tris-HCl, pH 8.8, 6.1 mM MgCl₂, 0.5 mM MnCl₂, 6.7 μ M EDTA, 10 mM β -mercaptoethanol, 1 mM primers, 1mM each dGTP, dTTP, and dCTP, 400 μ M dATP, and 2.5 units of *Taq* DNA polymerase (Promega, Madison, WI). After seven cycles of PCR (94°C for 40 sec, 50°C for 40 sec, and 72°C for 40 sec), the PCR product was amplified a second time using the same primers and *Vent* DNA polymerase (New England Biolabs, Beverly, MA) under the same conditions for 25 cycles. The resultant PCR product was gel purified, digested with *Bgl*II and

- 36 -

*Sac*II, and inserted into the *Bgl*II and *Sac*II sites of pVP16-CREB(*Bgl*II-*Sac*II)-LacZ (construction of which is described above). The resulting plasmids were transformed into DH5 α bacterial cells. Transformants were pooled and plasmid DNA was isolated by CsCl gradient centrifugation.

B. Construction and Use of pVP16-CREB(*Bgl*II-*Sac*II)-LacZ

A DNA fragment encoding the β -galactosidase gene was fused in frame to the carboxyl-terminal end of VP16-CREB as described above. The carboxy-terminal tag allowed identification of clones that contain frame-shift and nonsense mutations; colonies that remain positive for β -galactosidase were presumed to contain an open reading frame throughout the mutated region. To facilitate the subcloning of mutated sequences, a cassette encoding the CREB cDNA was generated that contained *Bgl*II and a *Sac*II sites flanking the 5' and 3' ends of the KID, respectively. These modifications altered the amino acid residue at position 168 from valine to alanine. The cDNA altered in this manner was indistinguishable from the original VP16-CREB and from VP16-CREB-LacZ when tested in the split hybrid assay. Primers complementary to regions flanking the KID were used in mutagenic PCR amplification reactions as described above under conditions which were optimized to achieve one to three mutations in the 177 bp region encoding the KID. PCR products were introduced into pVP16-CREB(*Bgl*II-*Sac*II)-LacZ in place of wild-type sequence. A library of mutated sequences was transformed into yeast assay strain Y1584 expressing LexA-CBD. Approximately 27,000 yeast transformants were screened, yielding about 5,000 colonies that were capable of growing on selective media supplemented with 10 μ g/ml of tetracycline and 1 mM of 3AT, determined as described below.

Two screening steps were performed to eliminate uninformative mutations and false positives. First, filter β -galactosidase assays were performed by standard methods [Vojtek, *et al.*, *Cell* 74:205-214 (1993)] on the 5,000 colonies which exhibited positive growth on media lacking

- 37 -

tryptophan, histidine, uracil, leucine, and lysine to eliminate expressed proteins having frame-shift and nonsense mutations. Five hundred thirty six colonies developed a dark blue color, whereas 412 colonies turned white and were presumed to express mutants containing either frame-shift or nonsense mutations. The other colonies developed a pale blue color, and control experiments suggested that these colonies may have expressed unstable lacZ fusion proteins. Pale blue colonies were not analyzed further.

DNA from 536 dark blue colonies was isolated and transformed into *E. coli* MC1066 cells. One hundred ninety three pVP16-CREB-(BgIII-SacII)-lacZ cDNAs were then isolated.

In a second screening step, the 193 cDNAs were separately re-transformed along with pLexA-CBD into the split-hybrid strain as well as into the two-hybrid L40 strain [Vojtek, *et al.*, *supra*] in order to identify false positives and confirm that the mutant CREB proteins did not interact with CBP. Among the 193 cDNAs re-screened, 152 did not interact with CBP in the yeast two-hybrid system, 15 interacted weakly, and 26 interacted like wild type CREB.

Following these two screening steps, the 152 CREB mutants were sequenced. Seventy CREB mutants were found to contain a single amino acid change. Sixty four CREB mutants contained two amino acid residue mutations and 13 mutants contained more than two amino acid mutations. Mutants containing more than one amino acid alteration were not analyzed further. The expression level of mutant proteins having one amino acid change were determined using a standard β -galactosidase assay.

The CREB mutations identified in the split-hybrid screen were shown to carry amino acid changes centered around the phosphorylation site at serine at position 133. No disrupting mutations were found to contain amino acid alterations outside of the region between amino acids 130 to 141. Most of the mutations abrogated the PKA phosphorylation region, but others were identified at isoleucine position 137, leucine at position 138, and leucine

- 38 -

at position 141. The mutations at positions 137, 138, and 141 generally changed the hydrophobic residues at these positions to polar residues. The ability of the split-hybrid system to detect only a limited number of CREB mutants, many of which have been proposed previously to disrupt CREB association with CBP [Parker, *et al.*, *Mol. Cell. Biol.* 16, 694-703, (1996)], indicates the specificity of the split-hybrid system.

These results lead to interesting suggestions. Various CREB mutations were identified which disrupt CREB-CBP interaction and the majority of disrupting mutations occurred in the CREB PKA phosphorylation motif. This result was consistent with previous observations that nonphosphorylated CREB and CBP do not interact [Kwok, *et al.*, *Nature* 370:223-226 (1994)]. The most common motif for PKA phosphorylation is an RXX(S/T)X amino acid sequence but RK(S/T)X and KRXX(S/T)X are also phosphorylated [Kemp and Pearson, *T.I.B.S.* 15, 342-346 (1990)]. The arginine residues in the phosphorylation site are critical for electrostatic interactions with acidic amino acid residues in the catalytic subunit of PKA [Knighon, *et al.*, *Science* 253, 414-420 (1991)], and consistent with this observation, CREB mutants with changes at arginine residues 130 and 131 were identified in the split hybrid assay that did not interact with CBP.

Results also showed that CREB mutations at amino acids proline at residue 132 and tyrosine 134 were unable to bind CBP. It is likely that the mutations at these residues adversely affect the structure of the phosphorylation motif, although these positions are generally thought to be less critical to CBP binding. It is possible that the substitution of proline at position 132 with threonine created a new phosphorylation site (RXTX) that interfered with the critical phosphorylation of serine at position 133. Although not generally thought to be part of the "classical" consensus PKA phosphorylation motif, hydrophobic amino acids are commonly found carboxy-terminal to PKA sites [Kemp, *et al.*, *T.I.B.S.* 19:440-444 (1994)]. The importance of these flanking residues may explain the frequent occurrence

- 39 -

of disrupting mutations involving tyrosine at position 134. Further studies will be directed to determining if mutations of proline at position 134 and tyrosine at position 134 directly disrupt phosphorylation of serine at position 133 or disrupt binding of CREB to CBP by some other mechanism.

5 In addition, substitution of serine at position 133 with threonine also prevented the interaction of CREB and CBP. PKA protein substrates containing a phosphorylatable threonine residue are known to exist in nature (*i.e.*, protein phosphatase inhibitor 1 and myelin basic protein), although they are less common than those with phosphorylatable serines [Zetterqvist, *et al.*, in *Peptides and Protein Phosphorylation*, (ed.) Kemp, B. E. (CRC Press, Boca Raton, FL), pp. 172-187 (1990)], and synthetic peptides containing serine to threonine substitutions are relatively poor substrates for PKA phosphorylation [Zetterqvist, *et al.*, *supra*]. In the split-hybrid assay, however, it is unclear whether the mutation of threonine at position 133 disrupts the CREB-CBP interaction or if the mutant fails to become phosphorylated. Despite previous observations that serine residue at position 133 of mammalian CREB can be phosphorylated by a variety of protein kinases other than PKA, for example calcium/calmodulin-dependent protein kinase II and IV, protein kinase C, and a nerve growth factor (NGF)-activated CREB kinase [Sheng, *et al.*, *Neuron* 4:571-582 (1990); Sheng, *et al.*, *Science* 252:1427-1430 (1991); Xie and Rothstein, *J. Immunol.* 154:1717-1723 (1995); Ginty, *et al.*, *Cell* 77:1-20 (1994)], it is not known which, if any, of these particular protein kinases are able to phosphorylate CREB at the serine at position 133 in yeast. The requirement for integrity of the entire RRXSX amino acid sequence, however, suggests that PKA is a reasonable candidate.

The second category of mutations were identified adjacent the PKA phosphorylation motif. Amino acids isoleucine at position 137 and leucine at position 138 have previously been suggested to be important for hydrophobic interactions of CREB with CBP [Parker, *et al.*, *Mol. Cell. Biol.* 16: 694-703 (1996)]. In this study, most of the mutations at position 137 and

- 40 -

138 converted these hydrophobic residues to polar amino acids. Thus, another possibility for the failure of these mutants to bind to CBP is that changes at these positions affect protein folding. Similarly, the mutation at position 141 substituted a polar residue for the wild-type hydrophobic leucine, and this mutation also has the potential to affect protein folding.

Substitution of the isoleucine at position 137 with a hydrophobic phenylalanine residue was found to disrupt the interaction between CREB and CBP as well. This result could have been the result of a detrimental effect on folding because of the steric hindrance associated with the comparatively larger size of phenylalanine. Alternatively, the proposed hydrophobic interactions between CREB and CBP are somewhat specific. Structural studies will be directed to definitively determine how these mutations affect binding.

Perhaps most surprising was the finding that critical mutations were restricted to a small region in the KID sequence, even though the relatively low affinity of phosphorylated CREB and CBP, determined to be between 250 and 400 nM by fluorescence anisotropy measurements [Kwok, *et al.*, *Nature* 370, 223-226 (1994)], is consistent with a restricted protein binding domain. The capability of the split-hybrid system to screen for a limited number of CREB mutants suggests that the system is highly specific, and thus, should be useful to identify mutations which disrupt interacts between other pairs of binding proteins.

- 41 -

Example 4 Tax/SRF Binding Interaction

To further investigate the feasibility of using the split-hybrid system to study protein-protein interactions, a pair of well characterized interacting proteins, SRF and Tax, was tested. Previous studies indicated that SRF and Tax interact in a standard yeast two-hybrid system suggesting that the proteins may be utilized in the split hybrid assay. Plasmid pLexA-SRF, containing a human SRF cDNA fused to the LexA DNA binding domain, was transformed into strain Y1584 along with either pVP16-Tax or pVP16 alone. As with the pLexA-VP16 transformation, the yeast strains co-expressing LexA-SRF and VP16-Tax failed to yield any colonies on medium lacking histidine. In contrast, when LexA-SRF was co-transformed with a vector encoding the VP16 activation domain alone, yeast growth occurred on medium lacking histidine, suggesting that TelR expression was not activated. These results demonstrated that a protein-protein interaction in the split-hybrid system can effectively prevent yeast growth and further indicated the utility of the assay for the study of various protein/protein interactions.

Example 5 Casein Kinase Binding Assays

In another example of use of the split hybrid assay to examine protein/protein interactions, Hrr25, a yeast casein kinase isoform, or human casein kinase I isoform δ , was employed in the assay with a known binding partner protein.

Previous work using the two hybrid assay had identified three genes encoding proteins which interact with the yeast casein kinase isoform Hrr25. Proteins encoded by the genes were designated TTH1, TTH2, and TTH3. The Hrr25 expression construct which was generated for use in the two hybrid assay was used in combination with the individual TTH encoding constructs in the split hybrid assay to determine if interaction between the

- 42 -

binding partners would decrease growth of assay yeast cells on media lacking histidine. Construction of the Hrr25 expression plasmid and isolation of plasmids encoding TTH proteins is discussed below.

In order to identify genes encoding proteins that interact with *S. cerevisiae* HRR25 CKI protein kinase, a plasmid library encoding fusions between the yeast GAL4 activation domain and *S. cerevisiae* genomic fragments ("prey" components) was screened for interaction with a DNA binding domain hybrid that contained the *E. coli* *lexA* gene fused to HRR25 ("bait" component). The fusions were constructed in plasmid pBTM116 which contains the yeast TRP1 gene, a *2 μ* origin of replication, and a yeast ADHI promoter driving expression of the *E. coli* *lexA* protein containing a DNA binding domain (amino acids 1 to 207).

Plasmid pBTM116::HRR25 encoding the *lexA*::HRR25 fusion protein was constructed in several steps. The DNA sequence encoding the initiating methionine and second amino acid of HRR25 was changed to a *Sma*I restriction site by site-directed mutagenesis using a Mutagenesis kit from BioRad (Richmond, California). The DNA sequence of HRR25 is set out in SEQ ID NO: 39. The oligonucleotide used for the mutagenesis is set forth below, wherein the *Sma*I site is underlined.

20 5'-CCTACTTAGGCGCGGCTTTTAATGATCC-3'
(SEQ ID NO: 37)

After digestion with *Sma*I, the resulting altered HRR25 gene was ligated into plasmid pBTM116 at the *Sma*I site to create the *lexA*::HRR25 fusion construct.

Interactions between bait and prey fusion proteins were detected in yeast reporter strain CTY10-5d (genotype=*MATa ade2 trp1-901 leu2-3,112 his 3-200 gal80 ura3::lexA op-lacZ*) [Luban, *et al.*, *Cell* 73:1067-1078 (1993)] carrying a *lexA* binding site that directs transcription of

- 43 -

lacZ. Strain CTY10-5d was first transformed with plasmid pBTM116::HRR25 by lithium acetate-mediated transformation [Ito, *et al.*, *J. Bacteriol.* 153:163-168 (1983)]. The resulting transformants were then transformed with a prey yeast genomic library prepared as GAL4 fusions in the plasmid pGAD [Chien, *et al.*, *Proc. Natl. Acad. Sci. (USA)* 21:9578-9582 (1991)] in order to screen the expressed proteins from the library for interaction with HRR25. A total of 500,000 double transformants were assayed for β -galactosidase expression by replica plating onto nitrocellulose filters, lysing the replicated colonies by quick-freezing the filters in liquid nitrogen, and incubating the lysed colonies with the blue chromogenic substrate 5-bromo-4-chloro-3-indolyl- β -D-galactoside (X-gal). β -galactosidase activity was measured using Z buffer (0.06 M Na_2HPO_4 , 0.04 M NaH_2PO_4 , 0.01 M KCl, 0.001 M MgSO_4 , 0.05 M β -mercaptoethanol) containing X-gal at a concentration of 0.002% [Guarente, *Meth. Enzymol.* 107:181-191 (1983)]. Reactions were terminated by floating the filters on 1M Na_2CO_3 and positive colonies were identified by their dark blue color.

Library fusion plasmids (prey constructs) that conferred blue color to the reporter strain co-dependent upon the presence of the HRR25/DNA binding domain fusion protein partner (bait construct) were identified. The sequence adjacent to the fusion site in each library plasmid was determined by extending DNA sequence from the GAL4 region. The sequencing primer utilized is set forth below.

5'-GGAATCACTACAGGATG-3' (SEQ ID NO: 38)

DNA sequence was obtained using a Sequenase version II kit (US Biochemicals, Cleveland, Ohio) or by automated DNA sequencing with an ABI373A sequencer (Applied Biosystems, Foster City, California).

Four library clones were identified and the proteins they encoded are designated herein as TH1 proteins 1 through 4 for Targets Interacting with

- 44 -

HRR25-like protein kinase isoforms. The TH11 portion of the TH1 clone insert corresponds to nucleotides 1528 to 2580 of SEQ ID NO: 40; the TH2 portion of the TH2 clone insert corresponds to nucleotides 2611 to 4053 of SEQ ID NO: 41; and the TH3 portion of the TH3 clone insert corresponds to nucleotides 248 to 696 of SEQ ID NO: 42. Based on DNA sequence analysis of the TH genes, it was determined that TH1 and TH3 were novel sequences that were not representative of any protein motif present in the GenBank database (July 8, 1993). TH2 sequences were identified in the database as similar to a yeast open reading frame having no identified function. (GenBank Accession No. Z23261, open reading frame YBL0506)

When the various TH proteins were used in the split hybrid assay in combination with Hrr25, it was observed that Hrr25/TH3 binding, previously determined to be weaker than Hrr25/TH2 or Hrr25/TH1 interactions, produced the lowest level of growth in the transformed yeast strain.

CK1 δ

In order to isolate cDNAs which encode proteins that interact with CK1 δ , the two hybrid assay was performed using a LexA-CK1 δ fusion protein as bait. The coding region of CK1 δ was subcloned into a *Bam*HI site of pBTM116 and transformed into a yeast strain designated CK1 δ /L40 (MAT a his3 Δ 200 trp1-901 leu2-3 112 ade2 LYS::(*lexAop*)₄HIS3 URA3::(*lexAop*)₈-*lacZ* GAL 4). CK1 δ /L40 was subjected to a large scale transformation with a cDNA library made from mouse embryos staged at days 9.5 and 10.5. Approximately 40 million transformants were obtained. Eighty-eight million were plated onto selective media lacking leucine, tryptophan and histidine. The ability of yeast transformants to grow in the absence of histidine suggested that there was an interaction between CK1 δ and some library protein.

In a second screening, interaction of the two proteins was

- 45 -

assayed by the ability of the interaction to activate transcription of β -galactosidase. Colonies that turned blue in the presence of X-gal were streaked onto media lacking leucine, tryptophan and histidine, grown up in liquid culture and pooled for isolation of total DNA. Isolated DNA was used to transform *E. coli* strain 600 which lacks the ability to grow on media lacking leucine. Colonies that grew were used for plasmid preparation and three classes of cDNA were identified. One class was closely related to a *Drosophila* transcription factor dCREBA.

When CKI δ /CREB interaction was examined in the split hybrid assay, cells were shown to grow on media containing histidine, but in the absence of histidine, growth was inhibited. Addition of small amounts of tetracycline to the cell culture restored the cell's ability to grow, suggesting that the interaction between CKI δ and CREB was very weak.

Example 6

AKAP 79 Binding Assays

Expression Plasmid Utilized

In still another example of use of the split hybrid assay to examine protein/protein interactions, an anchoring protein for the cAMP dependent protein kinase, AKAP 79, was utilized separately with binding partner proteins including the cAMP protein kinase regulatory subunit type I (RI), the cAMP dependent protein kinase regulatory subunit type II (RII) or calcineurin (CaN). Plasmids used in the assay were constructed as described below.

A 1.3 kb *NcoI/BamHI* fragment containing the coding region of AKAP 79 was isolated from a pET11d backbone and ligated into plasmid pAS1. Plasmid pAS1 is a 2 micron based plasmid with an ADH promoter linked to the Gal4 DNA binding subunit [amino acids 1-147 as described in Keegan et al., *Science*, 231:699-704 (1986)], followed by a hemagglutinin (HA) tag, polyclonal site and an ADH terminator. The expressed protein was

- 46 -

therefore a fusion between AKAP 79 and the DNA binding domain of Gal4.

Plasmids encoding RI, RII or CaN were isolated from a PACT murine T cell library in a standard two hybrid assay using the AKAP 79 expression construct described above. Plasmid PACT is a leu2, 2 micron based plasmid containing an ADH promoter and terminator with the Gal4 transcription activation domain II [amino acids 768-881 as described in Ma and Pastine, *Cell*, 48:847-853 (1987)], followed by a multiple cloning site. RI, RII and CaN encoding plasmids were isolated as described below.

A 500 ml SC-Trip yeast cell culture (OD₆₀₀ = 0.6-0.8) was harvested, washed with 100 ml distilled water, and repelleted. The pellet was brought up in 50 ml LISORB (100 mM lithium acetate, 10 mM Tris pH8, 1 mM EDTA pH8, and 1 M Sorbitol), transferred to a 1 liter flask and shaken at 220 rpm during an incubation of 30 minutes at 30°C. The cells were pelleted, resuspended in 625 μ l LISORB, and held on ice while preparing the DNA.

The DNA was prepared for transformation by boiling 400 μ l 10 mg/ml salmon sperm DNA for 10 minutes after which 500 μ l LISORB was added and the solution allowed to slowly cool to room temperature. DNA from a Mu T cell library was added (40-50 μ g) from a 1 mg/ml stock. The iced yeast cell culture was dispensed into 10 Eppendorf tubes with 120 μ l of prepared DNA. The tubes were incubated at 30°C with shaking at 220 RPM. After 30 minutes, 900 μ l of 40% PEG₃₃₅₀ in 100 mM Li acetate, 10 mM Tris, pH 8, and 1 mM EDTA, pH 8, was mixed with each culture and incubation continued for an additional 30 minutes. The samples were pooled and a small aliquot (5 μ l) was removed to test for transformation efficiency and plated on SC-Leu-Trip plates. The remainder of the cells were added to 100 ml SC-Leu-Trip-His media and grown for one hour at 30°C with shaking at 220 RPMs. Harvested cells were resuspended in 5.5 ml SC-Leu-Trip-His containing 50 mM 3AT (3-amino triazole) media and 300 μ l aliquots plated on 150 mm SC-Leu-Trip-His also containing 50mM 3AT. Cell were left to

- 47 -

grow for one week at 30°C.

After four days, titer plates were counted and 1.1×10^5 colonies were screened. Large scale β -gal assays were performed on library plates and ten positive clones were isolated for single colonies. One of these colonies grew substantially larger than the rest, and was termed clone 11.1. Sequence from clone 11.1 revealed an open reading frame 487 aa long which was correctly fused to the Gal-4 activation domain of pACT. The NIH sequence database was searched and the sequence was found to be closely homologous to the human calmodulin dependent protein phosphatase, calcineurin.

Additional screening using pACT Mu T-cell library DNA and the pAS1 AKAP 79 bait strain was performed in order to identify other AKAP 79 binding proteins by the protocol described above. Results from screening approximately 211,000 colonies gave one positive clone designated PACT2-1. Sequencing and a subsequent data base search indicated that the clone had 91% identity with rat type Ia regulatory subunit of protein kinase A (RI).

The library was rescreened using the same AKAP 79 bait and fifteen positives were detected from approximately 520,000 transformants. Of these fifteen, eleven were found to be homologous to the rat regulatory subunit type I of PKA. Each of these isolates were fused to the 5' untranslated region of RI and remained open through the initiating methionine.

Split Hybrid Analysis

In split hybrid analysis of AKAP79 binding interactions, a plasmid was first constructed for expression of a LexA::AKAP 79 fusion protein. An AKAP 79 coding region was excised from pAS AKAP 79 as an *NcoI/BamHI* fragment and inserted into pBTM116 previously digested with the same enzymes. The resulting plasmid was designated pBTM116-AKAP79. Approximately 50,000 W303 yeast cells (strain Y1665, see Table 1) in logarithmic growth were rinsed in media lacking histidine, suspended in 100 μ l to 200 μ l of the same media, and plated on agar lacking

- 48 -

histidine (to select for absence of protein/protein interaction) and also lacking leucine and tryptophan (to select for transformants bearing expression constructs encoding AKAP 79 and its binding partner). When RI was employed as the AKAP 79 binding partner, 2 to 4 μ M tetracycline and 5 mM 3AT were required to prevent the transformed host from growing under conditions where the expressed proteins interacted.

Once conditions were established under which growth of the transformed host was eliminated, various candidate inhibitor compounds were separately added to the agar. It was presumed that if one of the candidate compounds was capable of disrupting AKAP 79 interaction with the binding partner protein, growth of the transformed host should be detectable in the vicinity of the compound on the agar. In the split hybrid assay wherein AKAP 79 and RI binding was examined, 2 μ l of a 30 mM stock solution of ICOS Compound 4273 in DMSO, 2 μ l of a 10 mM stock solution of ICOS Compound 1062 in DMSO, and 2 μ l DMSO alone (as a negative control) were spotted on to the plate which was incubated at 30°C for four to five days. For ICOS Compound 4273 a ring of growth was detected.

In order to determine an IC_{50} for an inhibitor identified as described above, alternative methods may be used. In one method, the inhibitor compound is added to the agar over a range of concentrations. Ideally, the compound is diluted to the point that host cell growth is essentially not detectable.

In another method, a 96 well plate is used and the compounds of interest are serially diluted across one row of a 96 well plate, one compound per row. Media lacking histidine, tryptophan, and leucine is added (presuming that the expression plasmids encoding the binding partners also encode trp and leu proteins) along with the appropriately transformed host yeast strain. Tetracycline and 3AT are added at concentration previously determined to extinguish growth of the transformed host cell. After two to five days incubation at 30°C, the plate wells are read at approximately 600

- 49 -

nm using a plate reader. The concentration of inhibitor half way between zero and the lowest concentration that permits growth of the host cell to the level observed on media containing histidine is estimated to be IC₅₀.

A modification of this second method is particularly amenable for use in a high throughput screen of large numbers of candidate inhibitors. For example, rather than attempting to determine the IC₅₀ for a previously identified inhibitor, separate candidate inhibitors are added to each well of a 96 well plate, preferably at more than one concentration, and host cell growth determined after several days incubation. Inhibitory activity of compounds identified in this manner is confirmed on an agar plate and the IC₅₀ determined on 96 well plates, each assay as described above.

Example 7

General Application of The Split-Hybrid Screen

In order to examine general utility of the split hybrid system, various experiments were conducted with binding proteins known to interact. In addition, a number of control experiments were included in order to determine if the effects observed with the known binding partners were in fact due to protein/protein interaction.

A. Yeast Assay Strain Construction

Yeast transformants used in assays indicated below were derived from LYS2-deficient strains AMR69 (Mat a *his3 lys2 leu2 trp1*, URA3::LexA::lacZ) and AMR70 (Mat α *his3 lys2 trp1 leu2*, URA3::LexA::lacZ) [Hollenberg, et al., *Mol. Cell. Biol.* 15, 3813-3822 (1995); Chien, et al., *Proc. Natl. Acad. Sci. (USA)* 88:97578-9582 (1991); Fields and Song, *Nature* 340:245-246 (1989)]. Yeast were grown in YEED or selective minimal medium using standard conditions [Sherman, F., et al., *Methods in Yeast Genetics*, Cold Spring Harbor Lab., Cold Spring Harbor, NY (1986); *Methods in Enzymology*, Vol. 194 Guide to Yeast Genetics and

- 50 -

Molecular Biology. Eds. Christine and Fink]. Derivatives of both AMR69 and AMR70 strains lacking URA3 were first generated by streaking cells on synthetic media containing 5 mg/ml 5-Fluoro-orotic acid (5FOA) [Methods in Enzymology, Vol. 194 Guide to Yeast Genetics and Molecular Biology, Eds. Christine and Fink]. Two URA3 deficient mutants were required due to the fact that these strains were subsequently mated. URA3-deficient colonies were confirmed by testing for uracil auxotrophy and deletion of the URA::LexA::lacZ locus was confirmed by an absence of β -galactosidase activity assayed by standard methods. The mutant strains selected were designated 69-4 and 70-1.

Targeted integration of pRS306/8xLexAop/TetR was carried out by transforming [Hollenberg, et al., *Mol. Cell. Biol.* 15, 3813-3822 (1995)] the 69-4 strain with plasmid linearized at a unique NcoI site. The reporter gene construct was constructed using parental plasmid pRS306 which encodes URA3 as a selectable marker. Stably integrated plasmid thereby permitted selection on media lacking uracil. The positive uracil prototrophic strains were examined by Southern analysis to confirm insertion of the plasmid sequences.

Targeted integration of pRS303/2xetop-LYS was carried out by transformation [Hollenberg, et al., *supra*] of strain 70-1 with plasmid linearized at a unique HpaI site. The resulting lysine prototrophic strains were examined by Southern analysis to confirm insertion of the plasmid DNA.

The AMR69 derivative strain (MAT α) containing the pRS303/2xetop-LYS insertion was mated with the AMR70-derivative strain (MAT a) containing pRS306/8xLexAop/TetR and mated cells were selected on media lacking both lysine and uracil. Single colonies were grown up and tested for the ability to grow on media lacking histidine. The resulting strain was designated Y1584. In instances where yeast strains were transformed with other reporter gene pair combinations, the strains were uniquely designated. Yeast bearing integrated reporter gene constructs were

- 51 -

subsequently transformed [Hollenberg, *et al.*, *supra*] with plasmids encoding chimeric binding protein. Plasmids encoding the LexA DNA binding region were generally derived from parental plasmid pBTM116 which also encodes *TRP1* as a selectable marker. Plasmids encoding the VP16 transactivating domain were generally derived from parental plasmid pVP16 which also encodes *LEU2* as a selectable marker. Yeast cells which were successfully transformed with the four exogenous plasmids were therefore selected by an ability to grow on media lacking lysine, uracil, tryptophan, and leucine. Plasmids encoding various binding proteins were transformed into the yeast assay strain as indicated below.

B. Liquid Assay

After three days growth at 30°C on selection media as described above, a pool of colonies from each transformation was collected and diluted in 5 ml selective media. The mixture was vortexed and immediately sonicated for ten seconds. Cells in the resulting suspension were counted and seeded at 1000 cells/ml in selective media, 2 ml per 15 ml tube. Tetracycline, 3AT, and histidine were included as determined appropriate by the method described above. Each aliquot of cells was incubated with shaking for two days at 30°C and cell density measured at OD₆₀₀.

C. Characterization of the Assay

The utility of the split-hybrid assay was first determined using well characterized binding proteins and various controls.

In an initial study, Y1584 cells were transformed with plasmids pLexA-VP16 and pLeu. While the expressed proteins from the two plasmids do not interact, pLexA-VP16 encodes a fusion protein containing the VP16 activation domain fused directly to LexA which contains a DNA binding domain. The chimeric LexA-VP16 protein is a strong transactivator for a promoter containing LexA operators. Plasmid pLeu is essentially a blank used

- 52 -

as a control co-transformation plasmid.

Yeast transformed with the LexA-VP16 plasmid were able to express TetR protein as indicated by gel shift analysis using a *tet* operator oligonucleotide. In addition, the cells were unable to grow on media in the absence of histidine. Combined, these observations suggested that overexpressed TetR protein was capable of binding to *tet* operators and preventing the expression of *HIS3*. The transformed yeast grew on plates containing histidine, further indicating that overexpression of TetR did not have a toxic effect on the assay cells.

The results were consistent with previous observations and supported the earlier suggestion that activation of TetR expression, either through a single transcription factor or association of individual transcription factor domains, is capable of preventing assay cell growth on media lacking histidine, presumably by eliminating *HIS3* production.

Example 8

Split-Hybrid Assay With Weakly Interacting Binding Proteins

Protein/protein interaction was examined in the split-hybrid assay to determine utility of the system using two fusion proteins known to interact weakly. In this instance, the binding proteins were a 283 amino acid fragment of a cAMP regulatory binding protein (CREB283) fused to LexA and a fragment of the CREB binding protein consisting of the CREB binding domain (CBD) fused to VP16.

In this assay, yeast strain Y1584 described above was employed and transformation carried out as previously described. In a first assay, plasmids pLexA-CREB and pVP16-CBD were transformed into the cells and cell growth was observed in the absence of histidine in the media. Expression of the fusion proteins was confirmed by Western blotting. Attempts to decrease cell growth by titration with 3AT were unsuccessful in that the concentration of 3AT required to reduce growth in cells transformed with

- 53 -

plcxA-CREB and pVP16-CBD also eliminated growth in cells transformed with plcxA-CREB and the control plasmid pVP16.

In light of these results, two alternative approaches were taken in order to permit study of binding proteins wherein the interaction is relatively weak. Under the assumption that the system was failing at the level of TetR transcription, alternative approaches were taken in attempts to amplify the TetR effect on expression of *HIS3* gene. To achieve this end, assay cells were transformed with reporter constructs which encoded multiple *tet* operator sequences upstream from the *HIS3* gene. In the second approach, the *HIS3* promoter used to drive expression of the TetR gene was replaced with the stronger alcohol dehydrogenase (ADH) promoter.

In Y1596 cells wherein the ADH promoter replaced the *HIS3* promoter to drive TetR expression, transformation with plasmids plcxA-CREB and pVP16-CBD showed substantially decreased growth on his⁻ media as compared to that in assay strain Y1592 wherein the *HIS3* promoter was used to drive TetR expression. However, in cells transformed with plasmids plcxA-CREB 341-M1 and pVP16-CBD, no decrease in assay cell growth was detected on media lacking histidine. These results indicate that incorporation of the ADH promoter to drive TetR expression may be more useful in studies involving binding proteins that have low affinity.

When assay strains were utilized which incorporated plasmids wherein expression of the *HIS3* gene was driven by multiple copies of the *tet* operator, transformed cell lines did not grow well enough to indicate potential utility in subsequent assays.

- 54 -

Example 9 General Assay Methods

A. "Fine Tuning"

In instances where either of the test fusion proteins possesses intrinsic capacity for transcriptional activation, TetR will be expressed and growth of the assay strain media lacking histidine will be depressed proportional to the level of TetR expression. In order to restore growth of these cells to approximately the level observed on media containing histidine, the initially transformed assay yeast strains are grown in the presence of increasing concentrations of tetracycline which binds to the TetR gene product and prevents TetR binding to the *tet* operator. Precise titration of expressed TetR with tetracycline, only to the point that growth of the assay strain is restored to the level detected in the presence of histidine, permits detection of subsequent decreased growth of the assay strain following increased TetR expression resulting from interaction of the test binding proteins. The empirically determined tetracycline concentration is therefore employed to increase "signal-to-noise" ratios under assay conditions.

After an appropriate tetracycline concentration has been determined for each of the candidate assay strains, the cells are transformed with the second plasmid encoding the second fusion binding protein. As before, growth of each candidate assay strain is examined on media in the presence and absence of histidine. A desirable yeast assay strain is chosen which shows vigorous growth in the presence of histidine and negligible growth on media lacking histidine (indicative of the expected protein/protein interaction and resultant decreased expression of *HIS3*).

In instances where binding between the two test proteins is comparatively weak, TetR expression may not be sufficiently increased to abolish *HIS3* expression and cells expressing the resultant low levels of *HIS3* will still grow on media which lacks histidine. Cells which show this low level of viability are grown in the presence of increasing concentrations of 3-

- 55 -

aminotriazole (3AT), a competitive inhibitor in the histidine synthesis pathway, in order to reduce cell growth to negligible levels when plated on media lacking histidine. As with titration of TetR with tetracycline, addition of 3AT to the media is designed to increase the signal-to-noise ratio by providing significant changes in growth in the presence and absence of histidine in the media.

In a practical application of the methods for fine tuning, binding between CREB and the CREB binding protein (CBP) is illustrative. Growth of the yeast strain Y1584 transformed with pLexA-CBD, encoding the CREB binding domain (CBD) of CBP, and pVP16-CREB or pLexA-CBD and the control plasmid pVP16 was substantially decreased and virtually indistinguishable growth rates were detected in both instances on media lacking histidine. This observation indicated that the LexA-CBD protein product possessed sufficient transactivating capacity to eliminate HIS3 production. In order to distinguish growth differences between assay cells transformed with either pVP16 and pVP16-CREB, increasing amounts of tetracycline were added to the media lacking histidine.

In both transformants, tetracycline was able to relieve growth repression in a dose dependent manner, and at increasing concentrations of tetracycline, the difference in growth between the two colonies was increasingly magnified, with the most distinct growth difference observed following addition of tetracycline at 10 μ g/ml. Addition of tetracycline was therefore able to overcome the intrinsic transactivating capability of the LexA-CBD fusion protein.

Because the ultimate use of the split-hybrid system is for structure-function studies, mutagenesis studies, drug identification and library screens, it is important to minimize background growth that might be confused with disrupted protein-protein associations. This can be accomplished by the addition of 3AT, a competitive inhibitor of the HIS3 gene product. For instance, in the presence of 10 μ g/ml of tetracycline, the yeast strain

- 56 -

transformed with pLexA-CBD and pVP16-CREB still conferred approximately 12% growth of that observed in the presence of his⁺ media. To diminish this background, increasing concentrations of 3AT were added to the media in the presence of 10 μ g/ml of tetracycline. At the 3AT concentration of 0.25 mM, the growth of the yeast strain expressing LexA-CBD and VP16-CREB was below 5%, while the growth of the control strain was still maintained at 70% of control levels. These results indicate that split-hybrid system can be modulated by 3AT in addition to tetracycline in order to effectively increase the signal-to-noise ratio.

10 B. Preparation of yeast extracts

In order to assess the utility of various plasmids to function in the split-hybrid assay, a number of control experiments can be employed which lend insight into expression of a desired protein from the transformed plasmid. For example, standard immunological methodologies, i.e., immunoprecipitation, ELISA, etc., can be used to determine to the extent to which a desired protein is expressed. Similarly, a variation of the gel shift assay (discussed immediately hereafter) can be used to determine both if a protein is expressed and if the expressed protein is capable of DNA binding. In each of these control assays, a yeast extract is required which can be prepared as follows.

Extracts were prepared as described by Upaduri and Towle [*Mol. Cell. Biol.* 15:1499-1512 (1995)] and were used for electrophoretic mobility shift assays as discussed below. The yeast cells transformed with pLexA-VP16 were grown in 100 ml of selective synthetic medium lacking uracil, tryptophan, and lysine to a density of $A_{600} = 1$. Cells were harvested and washed with 5 ml of EB (containing 0.2 M Tris-HCl, pH 8.0, 400 mM (NH₄)₂SO₄, 10 mM MgCl₂, 1 mM EDTA, 10% glycerol, and 7 mM β -mercaptoethanol). Cells were transferred to microfuge tubes and collected by centrifugation. After resuspending in 200 μ l EB containing 1

- 57 -

- mM phenylmethylsulfonyl fluoride (PMSF), 1 μ g/ml leupeptin, and 1 μ g/ml pepstatin, a one-half volume of glass beads was added. The suspension was frozen in a -80°C freezer for 1 hour and thawed on ice. Thawed cells were vortexed at 4°C for 20 minutes, after which an additional 100 μ l BB was added, and cells were left on ice for 30 minutes. The suspension was centrifuged for 5 minutes, the supernatant was transferred to a new tube which was centrifuged for 1 hour in a microcentrifuge. The supernatant was then made to 40% with (NH₄)₂SO₄ and gently rocked for 30 minutes. After a 10 minute centrifugation, the pellet was resuspended in 300 μ l of 10 mM HEPES, pH 8.0, 5 mM EDTA, 7 mM β -mercaptoethanol, 1 mM PMSF, 1 μ g/ml leupeptin, and 1 μ g/ml pepstatin, and 20% glycerol. The resulting suspension was dialyzed against the same buffer, and aliquots were stored at -80°C.

C. Electrophoretic mobility shift assays

- Electrophoretic mobility shift assays were performed as described by Shih and Towle (*J. Biol. Chem.* 267:13222-13228 (1992)). Double-stranded *ret* operator oligonucleotides were prepared by combining equivalent amounts of complementary single-stranded DNA (SEQ ID NOS: 7 and 8) in a solution containing 50 mM Tris-HCl, pH 8.0, 10 mM MgCl₂, and 50 mM NaCl₂, heating the mixture to 70°C for 10 minutes, and then cooling to room temperature. The annealed oligonucleotides were labeled by filling in overhanging 5' ends using the Klenow fragment of *E. coli* DNA polymerase I with [α -³²P]dCTP. Binding reactions were carried out in 20 μ l containing 10 mM Tris-HCl, pH 7.5, 50 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol, 5% glycerol, and 2 mg of poly(dI-C). A typical reaction contained 20,000 cpm (0.5-1 ng) of end-labeled DNA with 3-5 μ g of yeast extract. Following incubation at 22°C for 30 minutes, samples were separated on a 4.5% nondenaturing polyacrylamide gel containing 50 mM Tris, 384 mM glycine, and 2 mM EDTA, pH 8.3. For competition binding experiments, the

- 58 -

conditions were exactly as above except that specific and nonspecific competitor DNAs were included in the binding mixture before the yeast extract was added. The concentration of tetracycline, a competitive inhibitor of TetR/*ret* operator binding, was 1 μ M when utilized.

Example 10

Application of the Split-Hybrid Assay to Identify Agents That Prevent Receptor Desensitization and Drug Tachyphylaxis

- Over half of the drugs that are used clinically affect the function of seven transmembrane receptors. Although many of the characteristics of these receptors are distinct, two general features appear to be conserved. One is the ability to signal through dissociation of heterotrimeric G proteins. The second is the capacity to lose responsiveness to ligand binding in a process termed desensitization which is mediated by receptor phosphorylation and the subsequent binding of factors that recognize the phosphorylated state of the receptor which prevents continued signaling. Desensitization results in an intrinsic limitation to drug action imposed by the action of the drug itself, *i.e.*, activation of a receptor by a hormone or drug initiates mechanisms that prevent subsequent responses to repeated administration of the same agent. The coupled mechanisms of activation and deactivation together have been termed "homologous desensitization," while the inability of a drug to maintain its efficacy is known as "tachyphylaxis." Even though the mechanisms underlying homologous desensitization have been worked out in great detail over the past few years, there are currently no useful pharmacological approaches available that prevent the inactivation mechanism.

- The potential clinical utility of agents that could prevent or modulate drug desensitization is enormous. Four examples where therapy is limited by the inability of receptors to maintain responsiveness to drugs include: (i) asthma wherein desensitization of airway adrenergic receptors

- 59 -

renders epinephrine treatment ineffective after a period of hours; (ii) congestive heart failure wherein desensitization of adrenergic and VTP receptors, coupled with an elevation of the β adrenergic receptor kinase (β ARK), prevents the inotropic effects of endogenous regulatory hormones; (iii) Parkinson's disease, wherein dopamine receptor desensitization limits the usefulness of agents like L-Dopa; and (iv) chronic pain wherein tolerance results from opiate receptor desensitization. Indeed, it is difficult to conceive of a pharmacological modality in use today that is not limited in its effectiveness by the phenomenon of desensitization.

The biochemical basis for G protein-coupled receptor desensitization involves three classes of proteins including arrestins, kinases and G-proteins, all of which have been cloned [Lefkowitz, *Nature Biotechnology* 14:283-286 (1996)]. Following activation of a seven transmembrane receptor, a region is phosphorylated by one or more G protein-coupled receptor kinases (known as GRKs 1-6). For example, in the β -adrenergic receptor (β AR) and rhodopsin, the cytoplasmic tail is phosphorylated [Premont, *et al.*, *J. Biol. Chem.* 269:6832-6841 (1994); Freedman, *et al.*, *J. Biol. Chem.* 270:17953-17961 (1995); Palczewski, *et al.*, *J. Biol. Chem.* 266:12949-12955 (1991); Palczewski, *et al.*, *J. Biol. Chem.* 270:15294-15298 (1995)] while in the m2 muscarinic receptor, the third cytoplasmic loop is phosphorylated [Nakata, *et al.*, *Eur. J. Biochem.* 220:29-36 (1994)]. The best characterized members of the family of G protein receptor kinases are the β AR kinase (β ARK) and rhodopsin kinase which are both membrane-associated. While rhodopsin kinase contains an intrinsic membrane targeting signal [Ingles, *et al.*, *Nature* 359:147-150 (1992)], β ARK appears to be targeted to the membrane by association with G protein $\beta\gamma$ subunits [Pitcher, *et al.*, *Science* 257:1264-1267 (1992); Ingles, *et al.*, *Nature* 359:147-150 (1992)]. Once the substrate receptor for each kinase is activated, presumably by ligand binding, the kinase associates and phosphorylates serine and threonine residues on the receptor. The phosphorylated receptor then becomes a binding target for one or more

- 60 -

other proteins. In the case of β AR, for example, phosphorylation allows binding of arrestin which prevents association with G proteins and promotes receptor sequestration and desensitization. Using the β AR as an exemplary desensitization model, it becomes apparent that multiple steps in the pathway appear to provide potential points of regulation each of which is amenable to the split-hybrid screen to identify molecules that can block the overall desensitization pathway. Specifically in the case of β AR, the split hybrid system can be used to identify small molecules that: (i) prevent interaction between β ARK and the G protein β subunit; (ii) inhibit β ARK activity; and (iii) disrupt the β ARK-arrestin complex.

A. Plasmid Constructions

The study of G-protein receptor kinases in the split-hybrid system involves three or more recombinant proteins or two or more recombinant proteins and a recombinant peptide library. In the split-hybrid system discussed above, two yeast primary expression plasmids are employed: pBTM116 [Barrel *et al.*, *Cellular Interactions in Development: a Practical Approach*, (ed) Hartley, IRL Press, Oxford, pp. 153-179 (1993)], which encodes the LexA-fusion protein and the TRP1 selectable marker, and pVP16 [Hollenberg *et al.*, *Mol. Cell. Biol.*, 15:3813-3822 (1995)], which encodes the VP16-fusion protein and the LEU2 selectable marker. In order to study interactions involving more than two recombinant proteins in the split-hybrid system, however, additional selectable markers are employed. Construction of additional yeast expression plasmids which are used to examine interactions between more than two binding proteins is discussed below.

- 61 -

1. Plasmid pDRM

A DNA fragment comprising the ADH promoter and LexA sites, the TeIR encoding gene, the nuclear localization signal, and the ADH terminator sequence are removed from pRS306/4xLexAop/ADH::TeIR with *SacI*, blunt-ended, and digested with *SacI*. The fragment is isolated and ligated into pRS303/2xtetop-L-Y32 which has previously been digested with *NotI*, blunt-ended, and digested with *SacI*. The resulting plasmid, designated pDRM, is integrated into the *LYS2* locus in the yeast genome as described above, and the resulting strain designated YTDRM. Placing the repressor gene and selectable marker reporter gene in the *LYS2* locus allows *URA3* to be used as a selectable marker.

2. Plasmid pRSURA3

A modified version of the pRS306 vector (Sikorski *et al.*, *Genetics*, **122**:19-27 (1989)) containing the *URA3* selectable marker gene is also used to encode additional recombinant proteins in the split-hybrid system. The plasmid, pRS426, has the 2 micron origin of replication inserted into a unique *AatI* site of pRS306. Plasmid pRS426 is further modified in the following manner:

(i) The ADH promoter sequence is amplified by PCR from BTM116 using primers which incorporate into the amplification product the DNA sequence encoding the SV40 large T antigen nuclear localization signal (NLS) and an initiating ATG sequence 3' to the ADH promoter. The ADH promoter/NLS/ATG sequence is inserted into the polylinker of pRS426.

(ii) The ADH terminator sequence is amplified by PCR from BTM116 using primers which incorporate into the product a DNA sequence encoding an antibody tag, for example, FLAG, hemagglutinin protein (HA), or thioredoxin (Trn) (FLAG, HA, and Trn antibodies are available through Santa Cruz Biotechnology, Santa Cruz, CA) and DNA sequences encoding stop codons in all three frames to the 5' end of the ADH terminator sequence.

- 62 -

The antibody tag/stop codon/ADH terminator sequence is inserted into the polylinker of pRS426.

3. Plasmid pRSAD22

PCR is used to engineer unique restriction sites, including for example, *BglII*, *Eco47II*, *MluI*, *NheI*, and *SphI*, immediately adjacent the 5' and 3' ends of the *URA3* cassette in pRSURA3. The *URA3* cassette is digested from pRSURA3 and replaced with the *ADE2* cassette which is amplified by PCR.

4. Plasmid pBTM116/AD4

A fragment containing the ADH promoter, polylinker, and ADH terminator is digested from pAD4 [Young *et al.*, *Proc. Nat'l. Acad. Sci. (USA)*, **86**:7989-7993 (1989)] with *BamHI*, blunt-ended and inserted into the blunt-ended *PvuI* site of BTM116 as described [Keegan *et al.*, *Oncogene*, **12**:1537-1544 (1996)], and the resulting vector designated pBTM116/AD4. PCR is also used to engineer a nuclear localization signal 3' of the ADH promoter as described above. This vector contains the *TRP1* selectable marker and can encode two recombinant proteins: (i) a LexA-fusion protein and (ii) a protein expressed from the pAD4 region of the vector.

B. β ARK and G Protein β Subunit Binding

In a first application of the split hybrid assay, disruption of binding between the carboxy-terminal domain of β ARK, containing the pleckstrin homology (PH) domain, and the G protein β subunit (G β_2) is examined. Previous work indicates that the PH domain of β ARK interacts directly with the $\beta\gamma$ subunits of G proteins [Fletcher, J.A., *et al.*, *Science* **257**:1264-1267 (1992) and Toulhara, K. *et al.*, *J. Biol. Chem.* **269**:10217-10220 (1994)]. Consistent with this observation is work by Punigilia, *et al.* [Punigilia, K.M., *et al.*, *J. Biol. Chem.* **270**:14251-14254 (1995)] which

indicates that $G\beta_2$ interacts with Raf1 in yeast and that the interaction is disrupted by βARK *in vitro*.

A DNA fragment containing the carboxy-terminal 222 amino acids (residues 467 to 689) of $\beta ARK1$, which includes the PH domain, is amplified by PCR from bovine $\beta ARK1$ [Pitcher *et al.*, *Science*, 257:1264-1267 (1992)] and the gel-purified amplification product is inserted into pBTM116. The resulting plasmid is designated LexA-COOH- βARK . A DNA fragment containing the entire coding sequence of $G\beta_2$ [Fong *et al.*, *Proc. Nat'l. Acad. Sci. (USA)*, 84:3792-3796 (1987)] is PCR amplified from pGEM-112f(-) $G\beta_2$ [Unigè-Luhfi *et al.*, *JBC*, 267:23409-23417 (1992)] and the gel-purified amplification product inserted into pVP16. The resulting plasmid is designated pVP16- $G\beta_2$. PCR is used in a similar manner to clone the carboxy-terminal domain of βARK into pVP16 and $G\beta_2$ into pBTM116.

βARK and $G\beta_2$ binding is first examined in the two-hybrid system to determine if expression of either binding partner as a fusion protein in yeast affects protein/protein interaction. Binding of the two proteins is then examined in the split hybrid assay in order to determine if protein/protein interaction is capable of abolishing growth of the assay yeast strain. As above, addition of tetracycline and/or 3-aminotriazole required to maximize the difference in growth in the presence and absence of histidine is empirically determined.

Split-hybrid yeast strains containing βARK and $G\beta_2$ subunits are used to screen libraries of small molecules. Several types of small molecule libraries can be examined in the split-hybrid assay, including for example, chemical libraries, libraries of products naturally produced by microorganisms, animals, plants and/or marine organisms, combinatorial, recombinatorial, peptidomimetic, multiparallel synthetic collection, protein, peptide and polypeptide libraries. A library of small peptides can be cloned into pRSTURA3 as described [Yang *et al.*, *Nuc. Acids Res.*, 23:1152-1156 (1995)] and Colas *et al.*, *Nature*, 380:548-550]. Peptides corresponding to the

carboxy-terminus of βARK or other GRKs which have previously been shown to block calcium channel desensitization in intact neurons, presumably by blocking βARK and $G\beta_2$ binding and subsequent trafficking of βARK to the cellular membrane [Diverse-Pierluissi, *et al.*, *Neuron* 16:579-585 (1996)] can be identified in such a screen. Further, it is important to show that the molecules identified through the split hybrid selection affect βARK - $G\beta$ interaction as opposed to, for example, tetracycline analogues identified in the screen that would not be useful to specifically modulate βARK / $G\beta_2$ binding.

B. Identification of βARK Inhibitors

In a second approach, agents that directly inhibit βARK function are identified in a modification of the split-hybrid system. While identification of specific βARK inhibitors may be difficult, preliminary data from split hybrid assays using CREB/CBP binding partners indicates that the system can be used to identify serine kinase inhibitors. The serine kinase results also suggest several approaches can be employed in attempts to overcome potential problems in identifying βARK inhibitors.

Briefly, binding between the phosphorylated G-protein coupled receptor (P-GR) and arresting is examined first in the standard two hybrid assay, followed by identification of inhibitors of P-GR/arresting binding in the split hybrid assay. For these studies, fragments of three G protein-coupled receptors are examined: the carboxy-terminal tail of β_2AR and the third cytoplasmic loop of the m2 muscarinic receptor. A DNA fragment containing the carboxy-terminal tail of the β_2AR (amino acids 330 to 413) is PCR amplified [Kolbika *et al.*, *JBC*, 262:7321-7327 (1987)] and the gel purified product inserted into pBTM116/Ad4 to produce a LexA- β_2AR fusion gene. The resulting plasmid is designated pBTM- β_2AR /AD4. A DNA fragment containing the third cytoplasmic loop of the human m2 muscarinic receptor (nucleotides 268-324) is amplified from pCEX-13m2 [Iaga *et al.*, *JBC*,

- 65 -

269:12594-12599 (1994)] by PCR and cloned into pBTM116/Ad4 creating a LexA-m2 fusion gene. The resulting plasmid is designated pBTM-m2/AD4. The entire bovine β ARK1 coding sequence [Benovic *et al.*, *Science*, 246:235-240 (1989)] is PCR amplified and cloned into the polylinker region originating from AD4 in pBTM- β -2AR/AD4 and pBTM-m2/AD4. The resulting plasmids are designated pBTM- β -2AR/AD4- β ARK and pBTM-m2/AD4- β ARK, respectively. PCR is used to amplify the DNA fragment containing bovine β arresting-1 (amino acids 1 to 437) [Lohse, *et al.*, *Science*, 248:1547-1550 (1990)]. This fragment is inserted into pVP16 and is designated pVP16- β arresting-1. PCR is used to amplify the DNA fragment containing rat β arresting-2 (amino acids 1 to 428) [Altamadal, *et al.*, *JBC*, 267:17882-17890 (1992)] which is inserted into pVP16 to give plasmid pVP16- β arresting-2. A PCR strategy is also used to clone β arresting into the pBTM116/AD4- β ARK plasmid and the β AR and m2 fragments into pVP16. As above, the yeast split-hybrid YIDRM strain is transformed with the P-GR- β arresting along with peptide libraries (cloned into pRSURA3) or grown following transformation in the presence of combinatorial drug libraries.

Inhibitors identified in the split hybrid assay should effect disruption of protein/protein interaction either by: (i) inhibiting β ARK phosphorylation of the receptor, thus preventing recognition of the receptor by β arresting, or (ii) by physical disruption of binding between the receptor and β arresting. Agents that allow yeast growth for trivial reasons, *i.e.*, tetracycline analogues, can be easily identified through use of simple controls.

A first potential problem to overcome in this study is that cytoplasmic β ARK enzyme must be targeted to the substrate receptor and, once targeted, must phosphorylate the receptor at appropriate sites. In normal cells, β γ association serves to target β ARK to the cell membrane; the β subunit binds to both the β ARK PH domain and the isoprenylated γ subunit in association with the membrane. One possible means to encourage the necessary specific interactions is to target the binding components in the assay

- 66 -

by tagging the proteins with nuclear localization signals, *i.e.*, β ARK, the receptor cytoplasmic tail, and β arresting, to the nucleus. The plasmids proposed for the study of the P-GR- β arresting interaction all contain nuclear localization signal sequences adjacent to recombinant gene sequence.

A second problem is somewhat more difficult to approach. The current model is that receptors must be activated by ligand binding before being phosphorylated by β ARK, *i.e.*, targeting of β ARK via β γ is not sufficient for receptor phosphorylation. There are two possible explanations for this requirement. The first is that phosphorylation sites on the receptor are masked in the absence of ligand and ligand binding causes a conformational change which "unmasks" the phosphorylation sites. If this is the case, a fragment of the receptor containing the immediate phosphorylation site may be used as the β ARK target. However, although peptides representing portions of the β AR cytoplasmic tail can be phosphorylated by β ARK, the K_m for the phosphorylation reaction is poor, suggesting that the kinase may require some other part of the receptor for binding and that the unmasking of this binding site by agonist is a critical step.

This problem is addressed in two ways. In the first, the m2 muscarinic receptor is used in place of the β AR in view of previous results which indicate that the m2 protein is a good substrate for β ARK. The third cytoplasmic loop of the m2 receptor serves as both the binding site and phosphorylation site for kinase and which should allow use of a LexA/m2 receptor third cytoplasmic loop fusion gene as one component in the screening system.

An alternative approach is to artificially mimic the activated state of the receptor. Haga, *et al.*, *J. Biol. Chem.*, 269:12594-12599 (1994)] have shown that the activity of β ARK can be stimulated *in vitro* in the presence of mastoparan, a bee venom peptide. Mastoparan is believed to mimic the cytoplasmic face of an activated receptor and has been shown to increase the affinity of β ARK for a GST-m2 receptor fusion protein by over

- 67 -

four orders of magnitude. The same effect can be seen by using peptides representing the flanking regions of the m2 third cytoplasmic loop. Thus, masoporan should also activate β ARK in the two-hybrid yeast strains, allow phosphorylation of the receptor fusion protein, and promote interaction with arresting. If masoporan is needed, oligonucleotides containing the coding and non-coding nucleotide sequences of the 14-mer peptide (INLKALALAKKIL-NH₂, SEQ ID NO: 43) are annealed and ligated into pSAD2. The yeast split-hybrid strain YIDRM is transformed with pBTM- β AR (or m2)/AD4- β ARK, pVP16-arresting, pRSAD2-masoporan, and a pRSURA3-peptide library or combinatorial drug library.

Numerous modifications and variations in the invention as set forth in the above illustrative examples are expected to occur to those skilled in the art. Consequently only such limitations as appear in the appended claims should be placed on the invention.

- 68 -

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Hoechst, Merl F.

(ii) TITLE OF INVENTION: Methods to Identify Compounds For Disrupting Protein/Protein Interactions

(iii) NUMBER OF SEQUENCES: 43

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSER: Marshall, O'Toole, Gerstein, Murray & Borun
(B) STREET: 6300 Sears Tower, 233 South Wacker Drive
(C) CITY: Chicago
(D) STATE: Illinois
(E) COUNTRY: United States of America
(F) ZIP: 60606-6402

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC DOS/MS DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER:
(B) FILING DATE:

(viii) ATTORNEY/AGENT INFORMATION:

(A) FIRM:
(B) REGISTRATION NUMBER:
(C) REFERENCE/DOCKET NUMBER: 7866/33424

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 312/474-6300
(B) TELEFAX: 312/474-0448
(C) TELEX: 25-3856

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

TTGTTGACG CTAGAGTCA CCGCAG

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 43 base pairs
(B) TYPE: nucleic acid

- 69 -

- (C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (11) MOLECULE TYPE: DNA
- (x1) SEQUENCE DESCRIPTION: SEQ ID NO:2:
TAACTCTAT CATTGATAGA GTAAATAGT AATGATATAT GCC
- (2) INFORMATION FOR SEQ ID NO:3:
- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (11) MOLECULE TYPE: DNA
- (x1) SEQUENCE DESCRIPTION: SEQ ID NO:3:
ATTACTCTAT CATTGATAGA GTAAATAGT AATGATATAT TC
- (2) INFORMATION FOR SEQ ID NO:4:
- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (11) MOLECULE TYPE: DNA
- (x1) SEQUENCE DESCRIPTION: SEQ ID NO:4:
AATTCGCTA GCCTCTGCA AGC
- (2) INFORMATION FOR SEQ ID NO:5:
- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (11) MOLECULE TYPE: DNA
- (x1) SEQUENCE DESCRIPTION: SEQ ID NO:5:
CGCAGCGTC GAAGAAATCA CATTACTTA TTAA
- (2) INFORMATION FOR SEQ ID NO:6:
- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (11) MOLECULE TYPE: DNA

- 70 -

- (x1) SEQUENCE DESCRIPTION: SEQ ID NO:6:
CGCAGCGCTA TACTAATAAA TGACGCGCA AG
- (2) INFORMATION FOR SEQ ID NO:7:
- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (11) MOLECULE TYPE: DNA
- (x1) SEQUENCE DESCRIPTION: SEQ ID NO:7:
CGCGTACTT ATCATGATA GAGTA
- (2) INFORMATION FOR SEQ ID NO:8:
- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (11) MOLECULE TYPE: DNA
- (x1) SEQUENCE DESCRIPTION: SEQ ID NO:8:
ATGAGATAGT AACTATCTCA TGGCG
- (2) INFORMATION FOR SEQ ID NO:9:
- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 48 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (11) MOLECULE TYPE: DNA
- (x1) SEQUENCE DESCRIPTION: SEQ ID NO:9:
CGCGTACTT ATCATGATA GAGCTAGAC TCTATCATAG ATAGAGTA
- (2) INFORMATION FOR SEQ ID NO:10:
- (1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
- (11) MOLECULE TYPE: DNA
- (x1) SEQUENCE DESCRIPTION: SEQ ID NO:10:
GGACGCGTG CATCGCGCT TCAAGATTC CTGGAG

- 71 -

- (2) INFORMATION FOR SEQ ID NO:11:
- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 36 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (11) MOLECULE TYPE: DNA
- (x1) SEQUENCE DESCRIPTION: SEQ ID NO:11:
GCCAGCGTG CATGCCACC GTACACGCT ACTCGA
- (2) INFORMATION FOR SEQ ID NO:12:
- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 34 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (11) MOLECULE TYPE: DNA
- (x1) SEQUENCE DESCRIPTION: SEQ ID NO:12:
CATGCATGC AAAAAAAG ATCATCCGC TAGG
- (2) INFORMATION FOR SEQ ID NO:13:
- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (11) MOLECULE TYPE: DNA
- (x1) SEQUENCE DESCRIPTION: SEQ ID NO:13:
CATGCATGC TTAGCATG GCATTATCAC AT
- (2) INFORMATION FOR SEQ ID NO:14:
- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (11) MOLECULE TYPE: DNA
- (x1) SEQUENCE DESCRIPTION: SEQ ID NO:14:
TATATGACT CATTATATG GG
- (2) INFORMATION FOR SEQ ID NO:15:
- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

- 72 -

- (11) MOLECULE TYPE: DNA
- (x1) SEQUENCE DESCRIPTION: SEQ ID NO:15:
TTTACACTT GCCTTGTT ATC
- (2) INFORMATION FOR SEQ ID NO:16:
- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 31 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (11) MOLECULE TYPE: DNA
- (x1) SEQUENCE DESCRIPTION: SEQ ID NO:16:
CGAGCCGAA GATGCTTGA TTAGATATAA G
- (2) INFORMATION FOR SEQ ID NO:17:
- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 49 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (11) MOLECULE TYPE: DNA
- (x1) SEQUENCE DESCRIPTION: SEQ ID NO:17:
CGGGATCCG CTTCTCTTC TTTTGGAG ACCCACTTC ACATTTAG
- (2) INFORMATION FOR SEQ ID NO:18:
- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (11) MOLECULE TYPE: DNA
- (x1) SEQUENCE DESCRIPTION: SEQ ID NO:18:
AATTCCTGA GTACTGTATG TACTACAGT AG
- (2) INFORMATION FOR SEQ ID NO:19:
- (1) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (11) MOLECULE TYPE: DNA
- (x1) SEQUENCE DESCRIPTION: SEQ ID NO:19:
AATTACTAG TATGTACATA CAGTACTGA GC

- 73 -

(2) INFORMATION FOR SEQ ID NO:20:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 32 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:20:
CCGAGATTC CAGACATAT CCAATCTTA TC

(2) INFORMATION FOR SEQ ID NO:21:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:21:
CCGAGATTC CTAATCCAT TATCATC

(2) INFORMATION FOR SEQ ID NO:22:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 32 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:22:
CATGCCATG CCAATCTAG ATTAATATA AG

(2) INFORMATION FOR SEQ ID NO:23:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:23:
GGGATTCGC CAGGACACA GAATCCACT

(2) INFORMATION FOR SEQ ID NO:24:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 32 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

- 74 -

(11) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:24:
CGGATTCCTG GCTGGTACC CAGATGCTT TG

(2) INFORMATION FOR SEQ ID NO:25:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:25:
CGGGATTCG GATGACCAT GACTGTGAG

(2) INFORMATION FOR SEQ ID NO:26:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:26:
CGGGATTCCT TAAATGACT TGTGGCATTA

(2) INFORMATION FOR SEQ ID NO:27:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 32 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:27:
CGGGATTCG CATGACCATG GAATCTGAG CC

(2) INFORMATION FOR SEQ ID NO:28:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

- 75 -

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:28:
CGCGATCG TGCTGCTT TCAGCAGCT G
31
(2) INFORMATION FOR SEQ ID NO:29:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(11) MOLECULE TYPE: DNA
(x1) SEQUENCE DESCRIPTION: SEQ ID NO:29:
ATGGTACCAG CGCGCGCTAG TCGTTTACA ACGTGTGAC
(2) INFORMATION FOR SEQ ID NO:30:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 38 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(11) MOLECULE TYPE: DNA
(x1) SEQUENCE DESCRIPTION: SEQ ID NO:30:
ATGGTACCAG CGCGCGCTAT TTTGACACC AGACCAAC
38
(2) INFORMATION FOR SEQ ID NO:31:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(11) MOLECULE TYPE: DNA
(x1) SEQUENCE DESCRIPTION: SEQ ID NO:31:
CGGAGATCTA AAGAGACTT TTCGCGAAC TCAG
34
(2) INFORMATION FOR SEQ ID NO:32:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(11) MOLECULE TYPE: DNA
(x1) SEQUENCE DESCRIPTION: SEQ ID NO:32:
CGGAGATCTT TACAGGAAG CTGAACTCT
29

- 76 -

(2) INFORMATION FOR SEQ ID NO:33:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(11) MOLECULE TYPE: DNA
(x1) SEQUENCE DESCRIPTION: SEQ ID NO:33:
CCACCGCGC AGTGCACAC CCAATTAC
29
(2) INFORMATION FOR SEQ ID NO:34:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(11) MOLECULE TYPE: DNA
(x1) SEQUENCE DESCRIPTION: SEQ ID NO:34:
CATCCGGCGT GGTGATGCA GCGGCTCA
28
(2) INFORMATION FOR SEQ ID NO:35:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(11) MOLECULE TYPE: DNA
(x1) SEQUENCE DESCRIPTION: SEQ ID NO:35:
GGCTATGAT ACGGCCCCC CAGACGAT
28
(2) INFORMATION FOR SEQ ID NO:36:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(11) MOLECULE TYPE: DNA
(x1) SEQUENCE DESCRIPTION: SEQ ID NO:36:
GGTATGAT CTACCCACG TACTCTC
28
(2) INFORMATION FOR SEQ ID NO:37:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 base pairs

- 77 -

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:37:
CCTACTCTTA GCGCCGGTC TTTTAATG ATCC
34

(2) INFORMATION FOR SEQ ID NO:38:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:38:
GGATCACTA CAGGATG
18

(2) INFORMATION FOR SEQ ID NO:39:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1485 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

- 78 -

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:39:
ATGACCTTAA GATGAGAG GAAATTGCT ATGAGCAGCA AGATTGGAG TGGTCTCTT
60
GGTGCATTT ACCAGCGAC GAATTAATT AGTGTGAG AGATGGCAT CAGCTGAAA
120
TGATCAGGT CGAGACATC TCAATTGAC TATGATCCC GCGTACAG ATACTATAG
180
GGTGTGTG GAATCCCGT CATTGATG TTGCGAGG AGGTGATA TAACTATG
240
GTATCGATC TTCTAGGCC ATCTTGGA GATTATCA ACTATGTA CAGAGGTC
300
TCCTTAAGA CCGTTATCAT GCTGCTTG CAATGTTT GCGTATCA GTATATCAT
360
CGAGGTGCT TCATTCATG AGATACAA CGAGCAACT TTTTATGG GTAGGACGC
420
CGGTGAGCA CGGTCAAT TATGATTC GGTATACA AGAATACCG AGATTCCAC
480
ACGATGTC ATATTCCTA CAGGAGAC AGTCTCTGA CAGTACGC TGTATGCA
540
AGTTCATA CGATCTTG ATTAGGCA ATGAGAGG ATACTTGA ATCTAGGT
600
TATGCTTGA TCTATTTT TAGGGTCT TTGCTAGGC AGGTTTGA AGCAACACC
660
AAGAAACAA AGTATGAG TATCAGGA AAGAAATTA AGTTAGCT GGAACCTA
720
TGTCAAGT TACCATGAG GTTCAGAA TATAGGCT ACTGTAGA TTGAATTC
780
GATGAGAC CAGATTATT GTCTTGCA AGCTGTTA AGATCTGAG TATTAACTA
840
GAGTATACA AGGACACTT GTTCATTTG ACAATGTGC GTTACACAA GCGATGTT
900
GAGAGCAA GGAATCTCT CATGAAAA GGTGATTTG AGGAAATG CATTGACGA
960
AATGCAGTA ACGACAGCA CACAGGCT GAACCTTCA ACAAATTA ACTTTAGCC
1020
ATGAGAAAT TCCCAACCA TTTCAGTAT TACGAGAG AAGACAACTA TATCTTCA
1080
CCAGAGAG TCAACAGCA AACTATCTT ATATATAT CAGCTCTTC TTACACAG
1140
GAATTATGA AGGACATGA TAAAGTATG GAAACTTGA GACACAGCA ACCAATGAC
1200
CAGTCCAAA GTTCGACGC ACACACAG CCCACAGC TACGACGA ACCAATGAC
1260
CAAGACCAA ATTATATCC TGACGTTA CTACACAGC AACAAAGAG TTCTCAGAG
1320
CAACGAGC AGTTCGAT GGTACACAC AGGCTACTC AGTATCCCC ACAATTAAC
1380
AGCATTAAT TTAATCTTA TCAAGCATC GTACTCCAC AATGATATC TATTCACAA
1440
CAGCCGCTC AAGATPACC AGCTGCCAG TCAATTGAT TGTAA
1485

(2) INFORMATION FOR SEQ ID NO:40:

(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2625 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

- 79 -

(1x) FEATURES:

(A) NAME/KEY: CDS
(B) LOCATION: 796..2580

(x1) SEQUENCE DESCRIPTION: SEQ ID NO:40:

CATTCTCTTA ATTCTTTAT GTGCTTTAC TACTTGTIT TACTTGAAC AATATGCTT
 60
 ATTCTTAGT ACTATGAT AAGACAGAA AAGAAAAA AAGAACAAAT AAGCTTAGCA
 120
 GAAATAGGT ATATTACT GTTACTTATA TACTTCAAG AAGATAGATT AATATGATG
 180
 CCGATTAGA AAAATATATA TAAAGGTCAAT CGATCCTCG CATTATATA TCCATATAA
 240
 GATACGAATC ACGGCAACT ATATTCAAG CTCATAGATA ATGCTGTATA GCGTCAACT
 300
 GCGAGAGAAA AGCTATTAAT TGAATACTAG CCGGTATGAA ACTGTGATTG ATTTACCTGG
 360
 GCTTACTCTA AAGAACATA AGTATATCTC ATGACAGAAAT CAAACACAAA TACAAATTT
 420
 ATCGAAACT CGGCGCACT GCGGCTGCC GGGAAAGGG AACACCGCTT CTATCTCGC
 480
 ACTATCTTCA TGGGCCAAT GGAAGCTAG ATATAGGAT TCCATTTAG CCGATAGAA
 540
 TGTAGGGTAA TACTGTGCG TATATAGTA TAGTATAGA ATTTATTAC CCGCGGAAA
 600
 TATTAGACA TCACTTAGCA CGAATTTAC GTCTAGAGAA AGTTAGATA TGGCCAAATA
 660
 ACCAGAGAAA ACAATATTG AATCTGTG AAGATTCGA CAGTTGTATA ATCTCTCTTA
 720
 AGCTCACTTA GTATCATAG TCTAATTAAT ATGCTTTGA ATCTGAAAAA AATTAAGTA
 780
 CCTTCGATT AAGCA ATG TCA CTG CCG CTA CCA CAC GCA TTG GAG AAC GTT
 831
 Met Ser Leu Pro Leu Arg His Ala Leu Glu Asn Val
 1 5 10
 ACT TCT GTT GAT AAG ATT TTA GAG GAC TTA TTA GTA CGT TTT ATT ATA
 879
 Thr Ser Val Asp Arg Ile Leu Glu Asp Leu Leu Val Arg Phe Ile Ile
 15 20 25
 AAT TGT CCG AAT GAA GAT TTA TCG AGT GTC GAG AGA GAG TTA TTT CAT
 927
 Asn Cys Pro Asn Glu Asp Leu Ser Ser Val Glu Arg Glu Leu Phe His
 30 35 40
 TTT GAA GAA GCC TCA TGG TTT TAC ACG GAT TTC ATC AAA TTG ATG AAT
 975
 Phe Glu Glu Ala Ser Trp Phe Tyr Thr Asp Phe Ile Lys Leu Met Asn
 45 50 55 60
 CCA ACT TTA CCC TCC CTA AAG ATT AAA TCA TTT GCT GAA TTG ATC ATA
 1023
 Pro Thr Leu Pro Ser Leu Lys Ile Lys Ser Phe Ala Glu Leu Ile Ile
 65 70 75
 AAA GTA TGT CCT CTG GTT TGG AAA TGG GAC ATA AGA GTG GAT GAG GCA
 1071
 Lys Leu Cys Pro Leu Val Trp Lys Trp Asp Ile Arg Val Asp Glu Ala
 80 85 90
 CTC CAG CAA TTC TCC AAG TAT AAG AAA AGT ATA CCG GTG AGG GGC GCT
 1119
 Leu Glu Gln Phe Ser Lys Tyr Lys Lys Ser Ile Pro Val Arg Gly Ala
 95 100 105
 GCC ATA TTT AAG GAG AAC CTG AGT AAA ATT TTA TTG GTA CAG GGT ACT
 1167
 Ala Ile Phe Asn Glu Asn Leu Ser Lys Ile Leu Leu Val Gln Gly Thr
 110 115 120

- 80 -

110 115 120
 GAA TCG GAT TCT TTG TCA TTC CCA AAG GGG AAG ATA TAT AAA GAT GAA
 1215
 Glu Ser Asp Ser Leu Ser Phe Pro Arg Gly Lys Ile Ser Lys Asp Glu
 125 130 135
 AAT GAC ATA GAT TGT TGC ATT AGA GAA GTG AAA GAA GAA ATT GGT TTC
 1263
 Asn Asp Ile Asp Cys Cys Ile Arg Glu Val Lys Glu Glu Ile Gly Phe
 145 150 155
 GAT TTG ACG GAC TAT ATT GAC GAC AAC CAA TTC ATT GAA AGA AAT ATT
 1311
 Asp Leu Thr Arg Tyr Ile Asp Asp Asn Gln Phe Ile Glu Arg Asn Ile
 160 165 170
 CAA GGT AAA AAT TAC AAA ATA TTT TTG ATA TCT GGT GTT TCA GAA GTC
 1359
 Gln Gly Lys Asn Tyr Lys Ile Phe Leu Ile Ser Gly Val Ser Glu Val
 175 180 185
 TTC AAT TTT AAA CCT CAA GTT AGA AAT GAA ATT GAT AAG ATA GAA TGG
 1407
 Phe Asn Phe Lys Pro Gln Val Arg Asn Glu Ile Asp Lys Ile Glu Trp
 190 195 200
 TTC GAT TTT AAG AAA ATT TCT AAA ACA ACG TAC AAA TAT ATC AAG
 1455
 Phe Asp Phe Lys Lys Ile Ser Lys Thr Met Tyr Ser Asn Ile Lys
 205 210 215
 TAT TAT CTG ATT AAT TCC ATG ATG AGA CCC TTA TCA ATG TGG TTA AAG
 1503
 Tyr Tyr Leu Ile Asn Ser Met Met Arg Pro Leu Ser Met Trp Leu Arg
 225 230 235
 CAT CAG AAG CAA ATA AAA AAT GAA GAT CAA TTG AAA TCC TAT GCG GAA
 1551
 His Gln Arg Gln Ile Lys Asn Glu Asp Gln Leu Lys Ser Tyr Ala Glu
 240 245 250
 GAA CAA TTG AAA TTG TTG TGG GGT ATC ACT AAG GAG GAG CAG ATT GAT
 1599
 Glu Gln His Lys Leu Leu Leu Gln Tyr Thr Lys Glu Glu Gln Ile Asp
 255 260 265
 CCC GGT AGA GAG TTG CTG AAT ATG TTA CAT ACT GCA GTG CAA GCT AAC
 1647
 Pro Gly Arg Glu Leu Leu Asn Met Leu His Thr Ala Val Gln Ala Asn
 270 275 280
 AGT AAT AAT AAT GCG GTC TCC AAC GGA CAG GTA CCC TCG AAC CAA GAG
 1695
 Ser Asn Asn Asn Ala Val Ser Asn Gly Gln Val Pro Ser Ser Gln Glu
 285 290 295
 CTT CAG CAT TTG AAA GAG CAA TCA GGA GAA CAC AAC CAA CAG AAG GAT
 1743
 Leu Gln His Leu Lys Glu Gln Ser Gly Glu His Asn Gln Gln Lys Asp
 305 310 315
 CAG CAG TCA TCG TTT TCT TCT CAA CAA CAA CCA CCT TCA ATA TTT CCA TCT
 1791
 Gln Gln Ser Ser Phe Ser Ser Gln Gln Gln Pro Ser Ile Phe Pro Ser
 320 325 330
 CTT TCT GAA CCG TTT GCT AAC AAT AAG AAT GTT ATA CCA CCT ACT ATG
 1839
 Leu Ser Glu Pro Phe Ala Asn Asn Lys Asn Val Ile Pro Pro Thr Met
 335 340 345
 CCA ATG GCT AAC GTA TTC ATG TCA AAT CCT CCA TTT TTT GCG AAG ATG
 1887
 Pro Met Ala Asn Val Phe Met Ser Asn Pro Gln Leu Phe Ala Thr Met
 350 355 360
 AAT GGC CAG CCT TTT GCA CCT TTC CCA TTT ATG TTA CCA TTA ACT AAC
 1935

- 81 -

Asn Gly Gln Pro Phe Ala Pro Phe Pro Phe Met Leu Pro Leu Thr Asn
365 370 375 380
AAT AGT AAT AGC GCT AAC CCT ATT CCA ACT CCG GTC CCC CCT AAT TTT
Asn Ser Asn Ser Ala Asn Pro Ile Pro Thr Val Pro Pro Asn Phe
385 390 395
AAT GCT CCT CCG AAT CCG ATG GCT TTT GCT GGT CCA AAC ATG CAT AAC
Asn Ala Pro Pro Asn Pro Met Met Ala Phe Gly Val Pro Asn Met His Asn
400 405 410
CTT TCT GGA CCA GCA GTA TCT CAA CCG TTT TCC TTG CCT CCT GCT CCT
Leu Ser Gly Pro Ala Val Ser Gln Pro Phe Ser Leu Pro Pro Ala Pro
415 420 425
TTA CCG AGG GAC TCT GGT TAC AGC AGC TCC TCC CCT GGG CAG TTG TTA
Leu Pro Arg Asp Ser Gly Tyr Ser Ser Ser Pro Gly Gln Leu Leu
430 435 440
GAT ATA CTA AAT TCG AAA AAG CCT GAC AGC AAC GTG CAA TCA ACC AAA
Asp Ile Leu Asn Ser Lys Lys Pro Asp Ser Asn Val Gln Ser Ser Lys
445 450 455
AAG CCA AAG CTT AAA ATC TTA CAG AGA GCA AGC GAC TTG AAT TCA CTC
Lys Pro Lys Leu Lys Ile Leu Gln Arg Gly Thr Asp Leu Asn Ser Leu
460 465 470 475
AAG CAA AAC AAT AAT GAT GAA ACT GCT CAT TCA AAC TCT CAA GCT TTG
Lys Gln Asn Asn Asn Asp Gln Thr Ala His Ser Asn Ser Gln Ala Leu
480 485 490
CTA GAT TTG TTG AAA AAA CCA ACA TCA TCG CAG AAG ATA CAC GCT TCC
Leu Asp Leu Leu Lys Lys Pro Thr Ser Ser Gln Lys Ile His Ala Ser
495 500 505
AAA CCA GAT ACT TCC TTT TTA CCA AAT GAC TCC GTA TCT GGT ATA CAA
Lys Pro Asp Thr Ser Phe Leu Pro Asn Asp Ser Val Ser Gly Ile Gln
510 515 520
GAT GCA GAA TAT GAA GAT TTC GAG AGT ACT TCA GAT GAA GAG GTG GAG
Asp Ala Gly Tyr Gly Asp Phe Gly Ser Ser Ser Asp Gly Val Gly
525 530 535 540
ACA GCT AGA GAT GAA AGA AAT TCA TTG AAT GTA GAT ATT GGG GTG AAC
Thr Ala Arg Asp Gly Arg Asn Ser Leu Asn Val Asp Ile Gly Val Asn
545 550 555
GTT ATG CCA AGC GAA AAA AAG GAC AGC CCA AGA AGT CAA AAG GAA AAA CCA
Val Met Pro Ser Gly Lys Asp Ser Arg Arg Ser Gln Lys Gly Val Pro
560 565 570
AAG AAC GAC GCA AGC AAA ACA ACA AAT TTG AAC GCT TCT GCA GAA TCT AAT
Arg Asn Asp Ala Ser Lys Thr Asn Leu Asn Ala Ser Ala Gly Ser Asn
575 580 585
ACT GTA GAA TGG GAG GCT GGG TAAATCTCA CCTCTCCACT TCAGAGTAC
Ser Val Gly Trp Gly Ala Gly
590 595
ACAGATCCA CAGTA

- 82 -

(2) INFORMATION FOR SEQ ID NO:41:
(1) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6854 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(11) MOLECULE TYPE: DNA
(1x) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 2050..4053
(1x1) SEQUENCE DESCRIPTION: SEQ ID NO:41:
AGCTCTCC TTTCCTTCA GTCTGTGAC TCCTGTCTT CCAGTAAAT GTTAAATTA 60
ATTGCAACT AGTTTGCAAT TGTACAAAC TCGCTTATC TTGTAAACA GAAGACGTA 120
TTTATATAT TGGGCTTAA TGTGTGAGT TTAATATAG ATAAAGTAG ACGAGTCT 180
GTCTTGTTT AACTATGAGG TTCAAGATA TAAGAGCGA GATAGAGAG TTAAGAAAA 240
AAAGGTTAG TTATATACG AAAGAAAGA AAGAGCGAA GTGCCACTA TAGCCCAAT 300
TCAAGAAATC AAGTCAGCAA AATACAGTAA TGTATAGAG ATACGCAAT GGTAAATCC 360
CTCAAGGCT CCGATACAGA AAAGCTAAG GAAATCTT AACTTACAG GCGTCAACA 420
GACTGAAAC AAGACTAAT TCTGTGAAA AGATGCTT AACTTGCTC TTGCATTAAC 480
TTTGAAAGAC AGGAACAAAG GTTATATGCG CTGATTAAC GTTGAAAGTA TATGATCTA 540
ATACACTTT GTTCTTAGG TCAATGCAAT ATGTTATCT CGAGAAAGG GTGCAGGCG 600
GTACCAATT ACTTGCCCT TTGGGTTAA ACAAGTTTA CATTTAAT ATATATAT 660
ATATGATGT GCGGTACCT ATATGCCCT CATACCAAT CATTTCTTG TTGCTGATG 720
GACTGCTTA TTTATTTCA AATGTAATT TTCAATTAT CTAGTCACT AAATTTGCA 780
AACTGCTTAC AATGTGCGT TAGCTGTGCG GTATCACTT CATTAACAG ATCGATTAAA 840
CTTTCAAGA AATTTGATC CTTGAATCC GCAAAATTC GATCTCACT TTGACCTCT 900
TGTAAAGTTC TTGCAGCAG GACTGCATCA GTAGAGCTA GCTACAAAG CCTTTTCTT 960
AGGAAGTAAT CCTTCAACT CCAATGGCT AATTAATTC CCAATGCTT CTGATTAAC 1020
TTGGAATATA TATCACTTC TTCAATATAT TGAACGTCA GAGCTTTAG ATCTGCAAT 1080
TTGATTAAC ACTTATTCA TAATCTAAC GACTGCTT GGGCATACA CTCACAGCG 1140
AGCTCATAC AATCTATAC TTTCATAGG TATAGCAAT CATTTTAG AATTTCTCA 1200
AGCTCAAACT TGAATTTAG ACTCTCCG AACTGCCCT TATAGTAAA AATTTGATA 1260
GCATTTCTA ATGAATCAC GGGTTTACA GAGTTTCA CCGTTTAAA GCATTTATA 1320
GCTCTACGT AGATATTCG TCGTTCGCT TCAATACAG CTTTCTTC ATATGACGA 1380

- 83 -

GCTTTCACAA AGAGTCTCC TGCCAGTTT AACCTCTTC TTAGACGGTA AATGGTGGCT 1440
 GCTTGACAC AAGATCAGC AGCTCTCA AACTGTATG AATCAGAAC GCTAACAAAT 1500
 TTCATGAAC CGATGAAAG AACACCTTC TTCACAGCT TAACACAGG GGAATATCA 1560
 ATTCCGGAT TTCATGTTA GTAAATGCC TTGCTAAT ACGAATCAC ATAGCTTCA 1620
 TTTTGTTCCT TTGATATTT TCCCTACTAC ATACTCTTT CAATACCTC ACGAGGCTCG 1680
 AATTTTAA CTTTCAGTT AATGATGTG TTCTTACTAT ATTCTCAGT CCGACAGAG 1740
 TTATTCAGA TAACTGCTT CGATGCTCC CACTCTTAT CATTACTCA ACTTACTCT 1800
 CCTATACTC GTGTGTCTT ATTAATTCGA GTTAATCGA GGTATAGAT TAGGTAAAC 1860
 TTCATGATG TCACGAACA CGATGCTGC AACTTGGCA TTTTCTCG GAAAGATTA 1920
 ACAAATTAAG CGAGCTTTC AGCTGAAAT ACCAGCAGGT CTTGGAGAT TAGCCACGA 1980
 AAGAGTCTA TATAGTACT ATAGAGGAG GCTACAGACT ACGGAAGCG TGTTCACAA 2040
 CAATGAA ATG GAG ACC AGT TCT TTT GAG AAT GCT CCT GCA GGC 2088
 Met Glu Thr Ser Phe Glu Asn Ala Pro Pro Ala Ala 10
 ATC AAT GAT GCT CAG GAT AAT AAT ATA AAT AGG GAG ACT AAT GAC CAG 2136
 Ile Asn Asp Ala Glu Asn Asn Ile Asn Thr Glu Thr Asn Asp Glu 25
 GAA ACA AAT CAG CAA TCT ATC GAA ACT AGA GAT GCA ATT GAC AAA GAA 2184
 Glu Thr Asn Glu Gln Ser Ile Glu Thr Arg Asp Ala Ile Asp Lys Glu 45
 AAT GGT GTG CAA AGC GAA ACT GGT GAG AAC TCT GCA AAA AAT GCC GAA 2232
 Asn Gly Val Glu Thr Glu Thr Gly Glu Asn Ser Ala Lys Asn Ala Glu 60
 CAA AAC GTT TCT TCT ACA AAT TTG AAT AAT GCC CCC ACC AAT GGT GCT 2280
 Glu Asn Val Ser Ser Thr Asn Leu Asn Asn Ala Pro Thr Asn Gly Ala 75
 TTG GAC GAT GAT GAT ATC CCA AAT GCT AAT GGT AAT AAA AAC AAT CCG 2328
 Leu Asp Asp Asp Val Ile Pro Asn Ala Ile Val Ile Lys Asn Ile Pro 85
 TTT GCT ATT AAA AAG CAA TTG TTA GAC ATT ATT GAA GAA ATG GAT 2376
 Phe Ala Ile Lys Lys Glu Gln Leu Leu Asp Ile Ile Glu Glu Met Asp 105
 CTT GCC CTT CCT TAT GGC TTC AAT TAC CAC TTT GAT AAC GGT ATT TTC 2424
 Leu Pro Leu Pro Tyr Ala Phe Asn Tyr His Phe Asp Asn Gly Ile Phe 115
 AAG GAA CTA GCC TTT GCG AAT TTC ACC ACT CCT GAA GAA ACT ACT CAA 2472
 Arg Gly Leu Ala Phe Ala Asn Phe Thr Thr Pro Glu Glu Thr Thr Gln 135
 GTG ATA ACT TCT TTG AAT GGA AAG GAA AAT AGC GGG AGG AAA TTG AAA 2520
 Val Ile Thr Ser Leu Asn Gly Lys Glu Ile Ser Gly Arg Lys Leu Lys 155

- 84 -

GTG GAA TAT AAA AAG ATG CTT CCC GAA GCT GAA AGA GAA AGA ATC GAG 2568
 Val Glu Tyr Lys Lys Met Leu Pro Gln Ala Glu Arg Glu Arg Ile Glu 160
 AGG GAG AAG AGA GAG AAA AGA GGA CAA TTA GAA GAA CAA CAC AGA TCG 2616
 Arg Glu Lys Arg Glu Lys Arg Gly Gln Leu Glu Glu Gln His Arg Ser 175
 TCA TCT AAT CTT TCT TTG GAT TCT TTA TCT AAA ATG AGT GGA AGC GGA 2664
 Ser Ser Asn Leu Ser Leu Asp Ser Leu Ser Lys Met Ser Gly Ser 190
 AAC AAT AAT ACT TCT AAC AAT CAA TTA TTC TCG ACT CTA ARG AAC GGC 2712
 Asn Asn Asn Thr Ser Asn Asn Gln Leu Phe Ser Thr Leu Met Asn Gly 210
 AAT AAT GCT AAT AGC ATG ATG AAC AGT AGT CCA ARG AAT AAT ACC ATT AAC 2760
 Ile Asn Ala Asn Ser Met Met Asn Ser Pro Met Asn Asn Thr Ile Asn 225
 AAT AAC AGT TCT AAT AAC AAC AAT AGT GGT AAC ATC ATT CTG AAC CAA 2808
 Asn Asn Ser Ser Asn Asn Asn Asn Ser Gly Asn Ile Ile Leu Asn Gln 240
 CTT TCA CTT TCT TCT GGC CAA CAT ACT TCT TCA TCG TTG TAC CAA ACA AAC 2856
 Pro Ser Leu Ser Ala Gln His Thr Ser Ser Ser 255
 GTT AAT AAT CAA GCC CAG ATG TCC ACT GAG AGA TTT TAT GGT CCT TTA 2904
 Val Asn Asn Gln Ala Gln Met Ser Thr Glu Arg Phe Tyr Ala Pro Leu 270
 CCA TCA ACT TCC ACT TTG CCT CTC CCA CCC CAA CAA CTG GAC TTC AAT 2952
 Pro Ser Thr Ser Thr Leu Pro Leu Pro Pro Gln Gln Leu Asp Phe Asn 290
 GAC CCT GAC ACT TTG GAA ATT TAT TCC CAA TTA TTG TTA TTT AAG GAT 3000
 Asp Pro Asp Thr Leu Glu Ile Tyr Ser Gln Leu Leu Phe Lys Asp 305
 AGA GAA AAG TAT TAT TAC GAG TTG GCT TAT CCC ATG GGT ATA TCC GCT 3048
 Arg Glu 320
 TTC CAC AAG AAG AAT ATC AAT GAT TTG TCG TCG TAC TTA GGG CTA GTA 3096
 Ser His Lys Arg Ile Ile Asn Val Leu Cys Ser 345
 GAA GTA TAT GAT CCA AGA TTT ATT ATT ATC AGA AGA AAG ATT CTG GAT 3144
 Glu Val Tyr Asp Pro Arg Phe Ile Ile Ile Arg Arg Lys Ile Leu Asp 350
 CAT GCT AAT TTA CAA TCT CAT TTG CAA CAA CAA GGT CAA ATG AAT TCT 3192
 His Ala Asn Leu Gln Ser His Leu Gln Gln Gly Gln Met Thr Ser 370
 GCT CAT CTT TTG GAG CCA AAC TCC ACT GGC GGC TCC ATG AAT AGC TCA 3240
 Ala His Pro Leu Gln Pro Asn Ser Thr Gly Gly Ser Met Asn Arg Ser 385
 CAA TCT TAT ACA AGT TTG TTA CAG GCC CAT GCA GCA GCT GCA GCG AAT 3288
 Gln Ser Tyr Thr Ser Leu Leu Gln Ala His Ala Ala 405

- 85 -

ACT ATT AGC AAT CAG GCC GTT AAC AAT TCT TCC AAC AGC AAT ACT ATT Ser Ile Ser Asn Gln Ala Val Asn Asn Ser Ser Asn Ser Asn Thr Ile 415 420 425	3336
AAC AGT AAT AAC GGT AAC GGT AAC AAT GTC ATC ATT AAT AAC AAT AGC Asn Ser Asn Asn Gly Asn Gly Asn Asn Val Ile Ile Asn Asn Asn Ser 430 435 440	3384
GCC AGC TCA ACA CCA AAA ATT TCT TCA CAG GGA CAA TTC TCC ATG CAA Ala Ser Ser Thr Pro Lys Ile Ser Ser Gln Gly Gln Phe Ser Met Gln 450 455 460	3432
CCA ACA CTA ACC TCA CCT AAA ATG AAC ATA CAC CMT AGT TCT CAA TAC Pro Thr Leu Thr Ser Pro Lys Met Asn Ile His His Ser Ser Gln Tyr 465 470 475	3480
AAT TCC GCA GAC CAA CGC CAA CAA CCT CCA CCA ACA CAG CAG CAA AAT Asn Ser Ala Asp Gln Pro Gln Gln Pro Gln Pro Gln Thr Gln Gln Asn 480 485 490	3528
GTT CAG TCA GCT GCG CAA CAA CAA CAA TCT TTT TTA AGA CAA CAA GCT Val Gln Ser Ala Ala Gln Gln Gln Gln Ser Phe Leu Arg Gln Gln Ala 495 500 505	3576
ACT TTA ACA CCA TCC TCA AGA ATT CCA TCC GGT TAT TCT GCC AAC CAT Thr Leu Thr Pro Ser Ser Arg Ile Pro Ser Gly Tyr Ser Ala Asn His 510 515 520	3624
TTT CAA ATC AAT TCC GTT AAT CCC TTA CTG AGA AAT TCT CAA ATT TCA Tyr Gln Ile Asn Ser Val Asn Pro Leu Ser Arg Asn Ser Gln Ile Ser 525 530 535	3672
CCT CCA AAT TCA CAA ATC CCA ATC AAC AGC CAA ACC CTA TCC CAA GCG Pro Pro Asn Ser Gln Ile Pro Ile Asn Ser Gln Thr Leu Ser Gln Ala 545 550 555	3720
CAA CCA CCA GCA CAG TCC CAA ACT CAA CAA CAG GTA CCA GTG GCA TAC Gln Asn Ala Ser Ser Gln Ser Gln Thr Gln Gln Arg Val Pro Val Ala Tyr 560 565 570	3768
CAA AAT GCT TCA TTG TCT TCC CAG CAG TTG TAC AAC CTT AAC GGC CCA Gln Asn GCT TCA TTG TCT TCC CAG CAG TTG TAC AAC CTT AAC GGC CCA 575 580 585	3816
TCT TCA GCA AAC TCA CAG TCC CAA CTG CTT CCA CAG CAC ACA AAT GGC Ser Ser Ala Asn Ser Gln Ser Gln Leu Leu Pro Gln His Thr Asn Gly 590 595 600	3864
TCA GTA CAT TCT AAT TTC TCA TAT CAG TCT TAT CAC GAT GAG TCC ATG Ser Val His Ser Asn Phe Ser Tyr Gln Ser Tyr His Asp Gln Ser Met 610 615 620	3912
TTG TCC GCA CAC AAT TTG AAT AGT GGC GAC TTG ATC TAT AAA TCT TTG Leu Ser Ala His Asn Leu Asn Ser Ala Asp Leu Ile Tyr Ser Leu 625 630 635	3960
AGT CAC TCT GGA CTA GAT GAT GGC TTG GAA CAG GGC TTG AAT COT TCT Ser His Ser Gly Leu Asp Asp Gly Leu Gln Gly Leu Asn Arg Ser 640 645 650	4008
TTA AGC GGA CTG GAT TTA CAA AAC CAA AAC AAG AAT CTA TGG Leu Ser Gly Leu Asp Leu Leu Gln Asn Lys Asn Leu Tyr 655 660	4053

- 86 -

TAAATATAC TTCCATTAAT CATATTAAT AGATTTT TGGATTTT ATATCCAG ATACAGTTAA TGAAGGTTC TTACACAGAA TAAATATTA AAAATATAT ATATATATA 4113 4117	4113
AAATCCATCA AAAACCAAT TGAATATAA TATATATAA AAAATATTA ACCAATATG 4223 4227	4223
AAATATTAAT TAATATCAT GATCAAGTTA ATTTTATG ACATACCTCA CTAATATG 4293 4297	4293
ATGAAAGAT GATATATAT GATATATG GCTACTTAA GTACAGCAAT GTATACAGC 4353 4357	4353
CAAAATATAC CCTCTTTT TTTTTCCT CTTTATGAT TTATTTTAA ACCATCACT 4413 4417	4413
TACTTTT TTTTATGCT TCTTTTCCA CATCTTTT TATATGAT TTATATATAC 4473 4477	4473
GATTTTAT CACAGATAT TTCTACTT ACTGATAT GTTTTTCAT TAATATATC 4533 4537	4533
CTTGCTCACT GCATCTATG CGATATTT ATATATATA AAGTGGGTA ACATATATC 4593 4597	4593
ATTATCATTT CTTTGTATC ATGATACAA GAAAGCATA AGTTTCATA GTCAAGCTC 4653 4657	4653
CAGGCAATTT GAACAAAGAA TTATCTTTT TTTTATCT TGAATTTCA ACTACATAC 4713 4717	4713
TATTTGTGTT GTTCTTGATT AATGGGGGC TTCAATTTT GAAATTAAGA GTCGGAAAA 4773 4777	4773
TAGCAGAAA ACAAGCATTA TTAATAGAG CAAAGAGAA AAGAACAT ATTAAGGTTA 4833 4837	4833
AAAAAGAAA AGCATGTCA TTCTTTCTC ATAGGTGTTA TTCAATCCG CCTCTCTTT 4893 4897	4893
CTTCTCTTT CATTAATTA TCTCCGATA ATTGCAAT AATTCATTA ACAAGCAACG 4953 4957	4953
AGCATCCG AAAACCCAAA AAAATGCT TATTAATAA CTTAGATATC GATGATCTGA 5013 5017	5013
TACATCTCA ATTATGCA AGCATACAA ATGACATAG AACACAGAA CAACTATTTCA 5073 5077	5073
ACTCCATGC CAGTATATCA CATGAAATTA GAGGATGTT TCGTCTGAT AAAGCCCTTA 5133 5137	5133
ATTAATATA AAAAGGTTT ATTCCCAAG AGTCAAACT TCGTCCAGAA AATTCCTTC 5193 5197	5193
AAATATCTT TAATAGGAC ATTAAGATAG GAGCATATG TTTCCTCAT ATTGAACCTTA 5253 5257	5253
ATATGAACT CACGAGGAG TTGAAGTTA GAATTAATTA TGAATCTATC AAATTTTCA 5313 5317	5313
ATTTGAAG ACTAATATCT AAATCTCAG TCAATACACC TTGAATTAAC AAAATATTA 5373 5377	5373
CATCATCCG TCTCTTAATC GAGTTTCAA GAAAGTTAA CAGGTTAAG CAACATAGG 5433 5437	5433
ATCTAGCAAC CAAATACAT GATACCAAT ATTCTCTGA TATATCCCA TTCAAGATTA 5493 5497	5493
AGCATCTTG GCATATGAT GTACATAG GCGAATCAT AAAAATAAATA AAAGATTTCA 5553 5557	5553
GTACAAATAG AACTTATCC ACTCGGAAG ATTAATATTA GTTCTTACA TTTTAAATA 5613 5617	5613
ACTATATGTC TGCATGATC ATTAATGTTA ACTATATCT TTGCATCAT ATTAATACG 5673 5677	5673
ATCAAGAAAA CAAATATAG AAAAATAAG AAGGAAAT ACCAAGATC AAAATTAAG 5733 5737	5733
TGATGTTTC CCAATATGA TACATTTTC TTCAATATA TGAATATTA CTTGTGATA 5793 5797	5793
CATCTCTTT TCGCAATC ATCAACAAA ATTAATGCT TTTATATCT GAGTTCTGC 5853 5857	5853
TAAATATAT GACTCTCA TATATTTT ATACACCA ACTAATAT CTAATTTTA 5913 5917	5913
AAGCCCCCG AAAATAAAT TATCGAAG GTAAAGCTTC TTTTGAAGA ATTATGCGC 5973 5977	5973

- 87 -

GGAGAGATGT CTGTATATGT CAATATCTT GCGTGTGCG AAAATGTTT ATCAATCTA 6033
 AGACTTGTA TTACGATGT GAACTGTTA TATCTATAT TCTACGGAG AAGAGAGATA 6093
 CAGAGAACCA TCCCTATCA AACGACGCCA AATTCATTT GATAGCTAT TTCTCCAGG 6153
 AAAATTTCA CTCGATATAC TATACCTTA GTTGATTTT AACTTACCC ATATACAGA 6213
 GGAAGGATA TGTCTAGTT TTATAGGAT TTCTATATT ATATCCAGA AAGAGCTCA 6273
 AATTGGAC TCCGAGAAA CCAATTCGG ATTAGGAT ATTACTTAC AGAACGTTT 6333
 GGAAGATATA ATGTGCTAA GTTGCTATTA AATTAAGAA CAGTCTAGA CTGTGATCA 6393
 ATATTAATAA TGAAGTACT TTTCAGCAG TTACCTTAA CGATATGCT AAATCAAGG 6453
 GATGATACC AACGACGTT GTGTGTGAT TGAACAACT TCAAGTTTG TATGCCATA 6513
 AACAGGCTC ATATTCAGT TTGATGAT TCACTATAT TATTAATC GATCTTGA 6573
 AAGGATTTA AATATTTAC AAACTTTGA GCTTAATAA CTATCTGCG GTCAATATG 6633
 ACAAATCTT GTTGGAACCT ATATATTAG GCGCTCAT TGTATTAAT GCGATATGA 6693
 ACTTGAACC CACCGCATTA GCGACGAGG CTCTTACAA TGAACATAT GCTCCGTAA 6753
 TTTCATATA CACGATATA GAAACTATA AAGAGTAG AGCACTATAT AAACGAGAA 6813
 GAGAGAGAG AAGAGTAGT GAGCAGAAA CATCAGCT T 6854

(2) INFORMATION FOR SEQ ID NO: 42:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2814 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(11) MOLECULE TYPE: DNA

(1x) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..696

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

GAA TTC CAA TAC ACC AAA CAG CTG CAT TTC CCT GTG GCG CCC AAA TTC 48
 GTC Phe Gln Tyr Thr Lys Gln Leu His Phe Pro Val Gly Pro Lys Ser 15
 1
 ACA AAC TGT GAG GTA GCG GAA ATT CTT TTA CAC TGC GAC TGG GAA AAG 96
 Thr Asn Cys Glu Val Ala Gln Ile Leu Leu His Cys Asp Trp Glu Arg 30
 20
 TAC ATA AAT GTT TTA AGT ATA ACA AGA ACA CCA AAT GTT CCT AGT GGT 144
 Tyr Ile Asn Val Leu Ser Ile Thr Arg Thr Pro Asn Val Pro Ser Gly 45
 35
 ACC AGT TTC ACG ACC AGA ACG ACG TAC ATG ATG TGC CGA TGG GAT GAC CAG 192
 Thr Ser Phe Ser Thr Arg Tyr Arg Met Phe Arg Trp Asp Asp Gln 60
 50

- 88 -

GGG CAA GGT TGC ATA TTA AAA ATA AGT TTT TGG GTG GAC TGG AAC GCA 240
 Gly Gln Gly Cys Ile Leu Lys Ile Ser Phe Trp Val Asp Trp Asn Ala 80
 65
 TCC AGT TGG ATC AAG CCA ATG GTA GAG ACG AAT TGT AAA AAT GAA CAA 288
 Ser Ser Trp Ile Lys Pro Met Val Glu Ser Asn Cys Lys Asn Gly Gln 95
 85
 ATT ACG GCC ACT AAG GAC TTG GTA AAG TTA GTC GAA GAA TTT GTA GAG 336
 Ile Ser Ala Thr Lys Asp Leu Val Lys Leu Val Glu Phe Val Glu 110
 100
 AAA TAC GTG GAA TTG ACG AAA GAA AAG GCA GAT ACA CTC AAG CCG TTG 384
 Lys Tyr Arg Gln Leu Ser Lys Lys Ala Asp Thr Leu Lys Pro Leu 125
 115
 CCC AGT GTT ACA TCT TTT GGA TCA CCT AGG AAA GTG GCA GCA CCG GAG 432
 Pro Ser Val Thr Ser Phe Gly Ser Pro Arg Lys Val Ala Pro Glu 140
 130
 CTG TCG ATG GTA CAG CCG GAG TCG AAA CCA GAA GCT GAG GCG GAA ATC 480
 Leu Ser Met Val Gln Pro Glu Ser Lys Pro Glu Ala Glu Ile 155
 145
 TCA GAA ATA GGC ACG GAC AAG TGG ACG TTT AAC TGG GTG AAC ATA ATA 528
 Ser Glu Ile Gly Ser Asp Arg Trp Arg Phe Asn Trp Val Asn Ile Ile 175
 165
 ATC TTG GTG CTC TTG GTG TTA AAT CTG CTG TAT TTA ATG AAG TTG AAC 576
 Ile Leu Val Leu Leu Val Leu Asn Leu Leu Tyr Leu Met Lys Leu Asn 185
 180
 AAG AAG ATG GAT AAG CTG ACG AAC CTC ATG ACC CAC AAG GAC GAA GTT 624
 Lys Lys Met Asp Lys Leu Thr Asn Leu Met Thr His Lys Asp Glu Val 205
 195
 GTA GCG CAC GCG ACT CTA TTG GAC ATA CCA GCC CAA GTA CAA TGG TCA 672
 Val Ala His Ala Thr Leu Leu Asp Ile Pro Ala Val Gln Trp Ser 215
 210
 AGA CCA AAG AAG GGA GAC GTG TTG TAAACAGATA ATCAATATAT ATTATATTA 726
 Arg Pro Arg Arg Gly Asp Val Leu 230
 225
 AGGTATGTA TGTTCGTATG GTATGAGAAA AAAAAAATAA AAGAGTCT ATTTGAGAA 786
 TGTAGAGGT GGTAGCTCG GATATTTGAG TCTGTAGGCT TCAATACCGA CAGTGCGT 846
 ACTGTAGAG GTTGCTCCCG TATTAAGTTC TCGTTTAA ATTTCTGGA AAAAAGAT 906
 TGAATGTTC AAGCTATCT CCGTGAAGA TTCTTCCCG AAGGTATTT GCTCCACTG 966
 CAGGAGATT GTGTTCGCT TTATTTTCG GTTGCTATA TTATTAACA GTATGTTG 1026
 CAATTTGCT TATTGAGAT CAATTTGGT GGTAGCTT TCAATATTT TCGAACCC 1086
 AAGCACTTC AACTTATA GTTGAGATA TTCTCTTTA TGAATTAAG ATAGAGCA 1146
 TGAAGAAC GCATTCATC GCAAACTC AAGGTGTCTA GATATTTGT TATCATATA 1206
 ATCACTAT CTAATATAT AATTAAGCA CAGGACAG TGTGTGCT GAGAGCA 1266
 CTGTGCTTT TCAATATAC CAATCATGT TTTCAGGATA ATTTCAATG GTATGCGG 1326

- 89 -

ATGAGAGCA	TAACTCTGA	CGCTCTAAC	CAATGCTTA	CTAGACTTCA	TGACTTACCG	1386
GGGCGAGTA	AAGATGCGA	AGGAAGAGAA	AAATGATATA	GAGGTGGTG	AATCGGAGAA	1448
ATAATTCGT	CCAACTATT	AATCAGAGCA	AAATATGTCT	TGTATGTTAT	TAAATATATC	1506
TATCTAACCA	TGTATTTACG	TATAAACTG	TCGATCTCA	TGCGCTAGCA	ATGAAAAAAT	1566
TTTTCTTAT	TTTTGTTAT	ATTCTCTT	GTTGCGTACT	TTTTTTTANT	GCGTTTCGCG	1628
GCAGAAAGCA	TTCCAGTTTA	CTGAGATGT	GTTATATCTA	AAAAATGTGA	TATGCTCTATT	1688
TTTGTTGTG	ATCTTTACTT	TACTGTATG	TACTGCTGA	GCGAGTACAC	TATGCTCTAT	1748
TTAGCGAGAC	GGATATGCGC	TCGGATACAG	CAGCGGCCCA	CGTATAGATA	GTTCTCTGTG	1806
TTTTAGAGTT	TGCCAAGATA	CTATCTAAAT	GCTCTCTCAG	CAGAGCTCTA	TTACTTTCTT	1866
TCCTCGBAAC	CAGCGATTCG	GTAATCTCTT	GTAACAGGTT	CAGAACTTG	TTCAATATGT	1926
AGCCCGACAG	AGCATATCAG	ACTTCTTGTT	TGTAAGGAGAC	ACCCCTCTAC	TGCTACTACTC	1986
TTCTTTGTG	CAGTTTGCCC	TTTAAATGCG	TTTTTACGCG	TATAGCATCT	TGCTATGTATT	2046
TGCGACAGTG	TATGCAATG	TGCTGACACA	GCGCCGATTT	GTTTCACTCC	ATGAGCTGAT	2106
TCAGAAAGTT	CTATATCTGAT	TGCTTGTGTT	ACCAATGCTT	ATATGATCAG	TGCACTATCTC	2166
GTTGTCTCT	TTTAGCTGTG	TATCTCTCA	CGCAATATCT	AGCCATGATG	CGCTATATGTT	2226
TATATCTAC	TGTCTGTGTT	GTTTATCTG	CTTTCATGTT	ACCTCATCTC	ATCTCCCGCT	2286
ACTTCTCTATA	TATTAAGAGCA	AAATATATTT	GCTTTTTCCT	CTCTCATGAT	AAAAAATTTT	2346
TCCCGAGGAT	ATAGAAAAAA	AAGAAATGTA	ATTATATATG	CGGTTATATTC	CTGCGGGTGC	2406
TTTTTTTAC	CTGTATGCTT	TTTCTCTCTG	TACATTTTTT	TTAATTTTTT	TTTGGGTTTT	2466
TTTTTTTTGA	TATTTTTCCT	TCCGAAACTGA	GTTACACAGA	TAACTCTGAC	TAGAGAAACT	2526
TTTTCTATCA	GAATATGTA	TGAGTTTGGT	TTCTCTAGAG	AATATGTTAT	AAAAAGATGT	2586
TGATGTGAG	GACATATATA	TACATCTTT	CGGCAATGCT	TTTTGAGATG	GAACTTTTCA	2646
ACTTATAGTT	ACGATATATC	AAATATCTAC	TTGCAGATG	CTAGAGAGCT	AGCTATATGT	2706
TACAGTATCC	AAAAGAGACA	GCTTATCCCA	AATCTATGAT	ACACAGAGAG	TATTTCAAGC	2766
TCCAGATCTA	GAATCTACGA	TTTGGGTAG	AAGATGCTTA	CGCTGCTAT		2814

(2) INFORMATION FOR SEQ ID NO:43:

(i) **SEQUENCE CHARACTERISTICS:**

(A) LENGTH: 14 amino acids

(D) TOPOLOGY: linear

(11) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

1 5 10
Ile Asn Leu Lys Ala Leu Ala Ala Leu Ala Lys Lys Ile Leu

- 06 -

WHAT IS CLAIMED IS:

1. A host cell transformed or transfected with DNA

a repressor gene encoding a repressor protein, said repressor gene under transcriptional control of a promoter;

a selectable marker gene encoding a selectable marker

control of an operator; said operator regulated by interaction with said repressor protein;

a first recombinant fusion protein gene encoding a first binding protein or binding fragment thereof in frame with either a DNA binding domain of a transcriptional activating protein or a transactivating domain of a transcriptional activating protein; and

a second recombinant fusion protein gene encoding a second binding protein or binding fragment thereof in frame with either a DNA binding domain of a transcriptional activating protein or a transactivating domain of a transcriptional activating protein, whichever domain is not encoded by the first fusion protein gene, said second binding protein or binding fragment thereof capable of interacting with said first binding protein or binding fragment thereof such that interaction of said second binding protein or binding fragment thereof and said first binding protein or binding fragment thereof brings into proximity a DNA binding domain and a transactivating domain forming a functional transcriptional activating protein; said functional transcriptional activating protein acting on said promoter to increase expression of said repressor gene.

- 91 -

2. The host cell of claims 1 wherein said DNA binding domain and said transactivating domain are derived from a common transcriptional activating protein.
3. The host cell of claim 1 wherein one or more of the repressor gene, the selectable marker gene, the first recombinant fusion protein gene, and the second recombinant fusion protein gene are encoded on distinct DNA expression constructs.
4. The host cell of claim 1 wherein said selectable marker protein is an enzyme in a pathway for synthesis of a nutritional requirement for said host cell such that expression of said selectable marker protein is required for growth of said host cell on media lacking said nutritional requirement.
5. The host cell of claim 1 wherein said host cell is a yeast cell or a mammalian.
6. The host cell of claim 2 wherein said selectable marker gene encodes HIS3;
7. The host cell of claim 2 wherein said repressor protein gene encodes a tetracycline resistance protein;
8. The host cell of claim 2 wherein said operator is a *tel* operator.
9. The host cell of claim 2 wherein said promoter is selected from the group consisting of the LexA promoter, the alcohol dehydrogenase promoter, the Gal4 promoter.

- 92 -

10. The host cell of claim 2 wherein said DNA binding domain derived from a protein selected from the group consisting of LexA and Gal4.
11. The host cells of claim 2 wherein said transactivating domain is derived from a protein selected from the group consisting of VP16 and Gal4.
12. The host cell of claim 2 wherein the first binding protein is CREB and the second binding protein is CBP.
13. The host cell of claim 2 wherein the first binding protein is Tax and the second binding protein is SRF.
14. The host cell of claim 2 wherein the first binding protein is casein kinase I and the second binding protein is CREB.
15. The host cell of claim 2 wherein the first binding protein is AKAP 79 and the second binding protein is selected from the group consisting of RI, RI β and calcineurin.
16. A method to identify an inhibitor of binding between a first binding protein or binding fragment thereof and a second binding protein or binding fragment thereof comprising the steps of:
 - a) growing host cells of any one of claims 1 through 15 in the absence of a test compound and under conditions which permit expression of said first binding protein or binding fragment thereof and said second binding protein or binding fragment thereof such that said first binding protein or fragment thereof and second binding protein or binding fragment thereof interact bringing

- 93 -

- into proximity said DNA binding domain and said transactivating domain forming said functional transcriptional activating protein; said transcriptional activating protein acting on said promoter to increase expression of said repressor protein; said repressor protein interacting with said operator such that said selectable marker protein is not expressed;
- b) confirming lack of expression of said selectable marker protein in said host cell;
- c) growing said host cells in the presence of a test compound; and
- d) comparing expression of said selectable marker protein in the presence and absence of said test compound wherein increased expression of said selectable marker protein is indicative that the test compound is an inhibitor of binding between said first binding protein or binding fragment thereof and said second binding protein or binding fragment thereof.

17. The method of claim 16 wherein the host cell is a yeast cell; the selectable marker gene encodes HIS3; transcription of the selectable marker gene is regulated by the *tet* operator; the repressor protein gene encodes the tetracycline resistance protein; transcription of the tetracycline resistance protein is regulated by the *LexA* promoter; the DNA binding domain is derived from *LexA*; and the transactivating domain is derived from VP16.

- 94 -

18. The method of claim 16 wherein the host cell is a yeast cell; the selectable marker gene encodes HIS3; transcription of the selectable marker gene is regulated by the *tet* operator; the repressor protein gene encodes the tetracycline resistance protein; transcription of the tetracycline resistance protein is regulated by the alcohol dehydrogenase promoter; the DNA binding domain is derived from *LexA*; and the transactivating domain is derived from VP16.
19. A kit to identify an inhibitor of binding between a first binding protein or binding fragment thereof and a second binding protein or binding fragment thereof, said inhibitor identified by the method of claim 16.

